

Binary Hypothesis Testing via Measure Transformed Quasi Likelihood Ratio Test

Nir Halay*, Koby Todros* and Alfred O. Hero†

*Ben-Gurion University of the Negev, †University of Michigan

Abstract

In this paper, the Gaussian quasi likelihood ratio test (GQLRT) for non-Bayesian binary hypothesis testing is generalized by applying a transform to the probability distribution of the data. The proposed generalization, called measure-transformed GQLRT (MT-GQLRT), selects a Gaussian probability model that best empirically fits a transformed probability measure of the data. By judicious choice of the transform we show that, unlike the GQLRT, the proposed test is resilient to outliers and involves higher-order statistical moments leading to significant mitigation of the model mismatch effect on the decision performance. Under some mild regularity conditions we show that the MT-GQLRT is consistent and its corresponding test statistic is asymptotically normal. A data driven procedure for optimal selection of the measure transformation parameters is developed that maximizes an empirical estimate of the asymptotic power given a fixed empirical asymptotic size. A Bayesian extension of the proposed MT-GQLRT is also developed that is based on selection of a Gaussian probability model that best empirically fits a transformed conditional probability distribution of the data. In the Bayesian MT-GQLRT the threshold and the measure transformation parameters are selected via joint minimization of the empirical asymptotic Bayes risk. The non-Bayesian and Bayesian MT-GQLRTs are applied to signal detection and classification, in simulation examples that illustrate their advantages over the standard GQLRT and other robust alternatives.

Index Terms

Hypothesis testing, higher-order statistics, probability measure transform, robust statistics, signal detection, signal classification.

This work was partially supported by United States-Israel Binational Science Foundation grant 2014334 and United States ARO MURI grant W911NF-11-1-0391.

I. INTRODUCTION

Classical binary hypothesis testing deals with deciding between two hypotheses based on a sequence of multivariate samples from an underlying probability distribution that is equal to one of two known probability measures [1]. When the probability distributions under each hypothesis are correctly specified the likelihood ratio test (LRT), which is the most powerful test for a given size [2], can be implemented. In many practical scenarios the probability distributions are unknown, even not up to some unknown parameters, and therefore, one must resort to suboptimal tests.

A popular suboptimal test of this kind is the Gaussian quasi LRT (GQLRT) [3]–[8] which assumes that the samples obey Gaussian distributions under each hypothesis. The GQLRT operates by selecting the Gaussian probability model that best fits the data. When the observations are i.i.d. this selection is carried out by comparing the empirical Kullback-Leibler divergences [9] between the underlying probability distribution and the assumed normal probability measures. The GQLRT has gained popularity due to its implementation simplicity, ease of performance analysis, and its geometrical interpretations. Despite the model mismatch, introduced by the normality assumption, the GQLRT has the appealing property of consistency when the mean vectors and covariance matrices are correctly specified and identifiable over the considered hypotheses [6]. However, in some circumstances, such as for certain types of non-Gaussian data, large deviation from normality can inflict poor decision performance. This can occur when the first and second-order statistical moments are weakly identifiable over the considered hypotheses, or in the case of heavy-tailed data when the non-robust sample mean and covariance provide poor estimates in the presence of outliers.

To overcome these limitations, several alternatives have been proposed in the literature. One straightforward approach is a non-Gaussian quasi LRT (NGQLRT) that involves more complex distributional models, e.g., elliptical, at the possible expense of increased implementation complexity, cumbersome performance analysis, and degraded performance under nominal Gaussian data. For example, by assuming Laplace distributed observations the NGQLRT for weak DC signal detection in additive i.i.d. noise is the well established sign detector [10], [11]. Although the sign detector is more resilient against heavy-tailed noise outliers as compared to the GQLRT, it has considerably poor performance when the noise is Gaussian [10]. Another approach is based on higher-order cumulants [12], [13] that may improve identifiability. However, unlike the first and second-order cumulants, used in the GQLRT, these quantities involve complicated tensor analysis [14]. Additionally, their empirical estimates are highly non-robust to outliers and have increased computational and sample complexity.

In this paper, a generalization of the GQLRT is proposed that operates by selecting a Gaussian

probability model that has the best empirical fit to a transformed probability distribution of the data. Under the proposed generalization, outlier-resilient tests can be obtained that involve higher-order statistical moments, and yet have the computational and implementation advantages of the standard GQLRT. This generalization, called the measure-transformed GQLRT (MT-GQLRT), is based on the measure transformation framework that was recently applied to canonical correlation analysis [15], [16], multiple signal classification (MUSIC) [17], [18] and parameter estimation [19], [20].

The measure transform is structured by a non-negative function, called the MT-function, and maps the probability distribution into a set of new probability measures on the observation space. By modifying the MT-function, classes of measure transformations can be obtained that have different useful properties that mitigate the model mismatch effect on the decision performance. Under the considered transform we redefine the measure-transformed (MT) mean vector and covariance matrix in the context of the hypothesis testing problem and show their relation to higher-order statistical moments. Furthermore, we reformulate the empirical estimates of the MT-mean and MT-covariance and restate the conditions on the MT-function for strong consistency and robustness to outliers. These quantities are then used to construct the proposed MT-GQLRT.

Similarly to the GQLRT, the proposed MT-GQLRT compares the empirical Kullback-Leibler divergences between probability distributions. The difference is that the MT-GQLRT compares the Kullback-Leibler divergences between the *transformed* probability distribution of the data and two normal probability measures that are characterized by the MT-mean vector and MT-covariance matrix under each hypothesis. Under some mild regularity conditions we show that the MT-GQLRT is consistent and its corresponding test statistic is asymptotically normal. Furthermore, given two training sequences from the probability distribution under each hypothesis, a data-driven procedure for optimal selection of the MT-function within some parametric class of functions is developed that maximizes an empirical estimate of the asymptotic power given a fixed empirical asymptotic size.

We go on to introduce a Bayesian extension of the proposed MT-GQLRT to mitigate the sensitivity of the standard Bayesian GQLRT [21]–[24] to model mismatch. In this context, it is worthwhile noting that other alternatives to the Bayesian GQLRT have been reported in [25]–[28] that apply linear and quadratic discriminant analysis to high-dimensional non-linear transformations of the observation vectors that map them into some reproducing kernel Hilbert spaces [29]. Although kernel methods can outperform the GQLRT, they may suffer from the following drawbacks. First, the high-dimensional mappings may have high computational complexity. Second, these methods are prone to over-fitting errors and require covariance regularization to avoid numerical instability.

Similarly to the non-Bayesian case, the Bayesian MT-GQLRT compares the empirical Kullback-Leibler divergences between a transformed conditional probability distribution of the data and two normal probability measures that are characterized by the MT-mean vector and MT-covariance matrix conditioned on each hypothesis. Like the non-Bayesian MT-GQLRT, the Bayesian MT-GQLRT can gain robustness against outliers under the same condition on the MT-function and its corresponding test-statistic is asymptotically normal. Furthermore, given two training sequences from the conditional probability distribution of each hypothesis, optimal selection of a parametric MT-function and the threshold value is carried out via joint minimization of the empirical asymptotic Bayes risk [1].

The proposed MT-GQLRT and its Bayesian extension are illustrated for signal detection and classification, respectively, in the presence of spherically contoured noise. By specifying the MT-function within the family of zero-centered Gaussian functions parameterized by a scale parameter, we show that the MT-GQLRT can significantly mitigate the model mismatch effect introduced by the normality assumption. More specifically, we show that the proposed MT-GQLRT outperforms the non-robust GQLRT and other robust alternatives and attains decision performance that are significantly closer to those obtained by the omniscient LRT that, unlike the proposed test, requires complete knowledge of the likelihood functions under each hypothesis.

Interestingly, in [30]–[32] it was shown that the performance of suboptimal binary hypothesis tests can be significantly improved by adding noise to the data that modifies the underlying probability distribution. We emphasize that here, we use a different probability measure-transformation approach that does not involve noise addition.

The paper is organized as follows. In Section II, the GQLRT is reviewed. Section III reviews the principles of the considered probability measure transform. In Section IV, we use this transformation to construct the non-Bayesian MT-GQLRT. Its extension for Bayesian hypothesis testing is developed in section V. The MT-GQLRT and its Bayesian extension are applied to signal detection and classification, respectively, in Section VI. In Section VII, the main points of this contribution are summarized. The proofs of the propositions and theorems stated throughout the paper are given in the Appendix.

II. GAUSSIAN QUASI LIKELIHOOD RATIO TEST: REVIEW

A. Preliminaries and problem formulation

We define the measure space $(\mathcal{X}, \mathcal{S}_{\mathcal{X}}, P_{\mathbf{X};H})$, where $\mathcal{X} \subseteq \mathbb{C}^p$ is the observation space of a complex-valued random vector \mathbf{X} , $\mathcal{S}_{\mathcal{X}}$ is a σ -algebra over \mathcal{X} and $P_{\mathbf{X};H}$ is a probability measure on $\mathcal{S}_{\mathcal{X}}$ parameterized by a hypothesis indicator non-random symbolic parameter H that belongs to a pair set $\mathcal{H} \triangleq \{H_0, H_1\}$. It

is assumed that the family $\{P_{\mathbf{X};H} : H \in \mathcal{H}\}$ is absolutely continuous w.r.t. a dominating σ -finite measure ρ on $\mathcal{S}_{\mathcal{X}}$, such that the Radon-Nikodym derivative [33]

$$f(\mathbf{x}; H) \triangleq \frac{dP_{\mathbf{X};H}(\mathbf{x})}{d\rho(\mathbf{x})},$$

exist for all $H \in \mathcal{H}$. The function $f(\cdot; H)$ is called the density function of \mathbf{X} , parameterized by H . Let $g : \mathcal{X} \rightarrow \mathbb{C}$ denote an integrable scalar function. The expectation of $g(\mathbf{X})$ under $P_{\mathbf{X};H}$ is defined as:

$$\mathbb{E}[g(\mathbf{X}); P_{\mathbf{X};H}] \triangleq \int_{\mathcal{X}} g(\mathbf{x}) dP_{\mathbf{X};H}(\mathbf{x}),$$

where $\mathbf{x} \in \mathcal{X}$.

Given a sequence of samples from $P_{\mathbf{X};H}$ we consider the problem of testing between the null and alternative hypotheses $H = H_0$ and $H = H_1$, respectively, when $P_{\mathbf{X};H_0}$ and $P_{\mathbf{X};H_1}$ are unknown.

B. The GQLRT

Given a sequence of samples from the underlying probability distribution $P_{\mathbf{X};H}$, the GQLRT operates by comparing the empirical Kullback-Leibler divergences between $P_{\mathbf{X};H}$ and two complex circular Gaussian probability measures [34], $\Phi_{\mathbf{X};H_k}$, $k = 0, 1$. These Gaussian measures are characterized by the mean vectors $\boldsymbol{\mu}_{\mathbf{X};H_k} \triangleq \mathbb{E}[\mathbf{X}; P_{\mathbf{X};H_k}]$, $k = 0, 1$, and the covariance matrices $\boldsymbol{\Sigma}_{\mathbf{X};H_k} \triangleq \mathbb{E}[\mathbf{X}\mathbf{X}^H; P_{\mathbf{X};H_k}] - \boldsymbol{\mu}_{\mathbf{X};H_k}\boldsymbol{\mu}_{\mathbf{X};H_k}^H$, $k = 0, 1$, that are assumed to be known (up to some redundant constants).

The Kullback-Leibler divergence (KLD) between $P_{\mathbf{X};H}$ and $\Phi_{\mathbf{X};H_k}$, $k \in \{0, 1\}$ is defined as [9]:

$$D_{\text{KL}}[P_{\mathbf{X};H}||\Phi_{\mathbf{X};H_k}] \triangleq \mathbb{E}\left[\log \frac{f(\mathbf{X}; H)}{\phi(\mathbf{X}; H_k)}; P_{\mathbf{X};H}\right], \quad (1)$$

where $f(\mathbf{x}; H)$ and

$$\phi(\mathbf{x}; H_k) \triangleq \det^{-1}[\pi\boldsymbol{\Sigma}_{\mathbf{X};H_k}] \exp\left(-(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X};H_k})^H (\boldsymbol{\Sigma}_{\mathbf{X};H_k})^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X};H_k})\right) \quad (2)$$

are the density functions of $P_{\mathbf{X};H}$ and $\Phi_{\mathbf{X};H_k}$ respectively, w.r.t. the dominating σ -finite measure ρ on $\mathcal{S}_{\mathcal{X}}$. By (1) and (2), one can verify that when $\phi(\cdot; H_0) \neq \phi(\cdot; H_1)$, the difference $D_{\text{KL}}[P_{\mathbf{X};H}||\Phi_{\mathbf{X};H_0}] - D_{\text{KL}}[P_{\mathbf{X};H}||\Phi_{\mathbf{X};H_1}]$ will be negative if $H = H_0$ and positive if $H = H_1$. This motivates the use of the empirical estimate of this difference as a test statistic.

Hence, given a sequence of samples \mathbf{X}_n , $n = 1, \dots, N$ from $P_{\mathbf{X};H}$, an empirical estimate of (1) is defined as:

$$\hat{D}_{\text{KL}}[P_{\mathbf{X};H}||\Phi_{\mathbf{X};H_k}] \triangleq \frac{1}{N} \sum_{n=1}^N \log \frac{f(\mathbf{X}_n; H)}{\phi(\mathbf{X}_n; H_k)}.$$

and the resulting test statistic, which is independent of the unknown density $f(\cdot; \cdot)$, takes the form:

$$\begin{aligned}
T &\triangleq \hat{D}_{\text{KL}}[P_{\mathbf{X};H}||\Phi_{\mathbf{X};H_0}] - \hat{D}_{\text{KL}}[P_{\mathbf{X};H}||\Phi_{\mathbf{X};H_1}] \\
&= \frac{1}{N} \sum_{n=1}^N \psi(\mathbf{X}_n) \\
&= \left(D_{\text{LD}} \left[\hat{\Sigma}_{\mathbf{X}} || \Sigma_{\mathbf{X};H_0} \right] + \|\hat{\mu}_{\mathbf{X}} - \mu_{\mathbf{X};H_0}\|_{(\Sigma_{\mathbf{X};H_0})^{-1}}^2 \right) \\
&- \left(D_{\text{LD}} \left[\hat{\Sigma}_{\mathbf{X}} || \Sigma_{\mathbf{X};H_1}^{(u)} \right] + \|\hat{\mu}_{\mathbf{X}} - \mu_{\mathbf{X};H_1}\|_{(\Sigma_{\mathbf{X};H_1})^{-1}}^2 \right),
\end{aligned} \tag{3}$$

where

$$\psi(\mathbf{X}) \triangleq \log \frac{\phi(\mathbf{X}; H_1)}{\phi(\mathbf{X}; H_0)},$$

is the log-ratio between the assumed Gaussian densities under H_1 and H_0 , $D_{\text{LD}}[\mathbf{A}||\mathbf{B}] \triangleq \text{tr}[\mathbf{A}\mathbf{B}^{-1}] - \log \det[\mathbf{A}\mathbf{B}^{-1}] - p$ is the log-determinant divergence [35] between positive definite matrices $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{p \times p}$, $\|\mathbf{a}\|_{\mathbf{C}} \triangleq \sqrt{\mathbf{a}^H \mathbf{C} \mathbf{a}}$ denotes the weighted Euclidean norm of a vector $\mathbf{a} \in \mathbb{C}^p$ with positive-definite weighting matrix $\mathbf{C} \in \mathbb{C}^{p \times p}$ and $\hat{\mu}_{\mathbf{X}} \triangleq \frac{1}{N} \sum_{n=1}^N \mathbf{X}_n$ and $\hat{\Sigma}_{\mathbf{X}} \triangleq \frac{1}{N} \sum_{n=1}^N \mathbf{X}_n \mathbf{X}_n^H - \hat{\mu}_{\mathbf{X}} \hat{\mu}_{\mathbf{X}}^H$ denote the standard sample mean vector (SMV) and sample covariance matrix (SCM). The decision rule based on the test statistic (3) is given by

$$T \underset{H_0}{\overset{H_1}{\gtrless}} t, \tag{4}$$

where $t \in \mathbb{R}$ denotes a threshold determined either to ensure a given level of false alarm probability or to minimize the (Bayes) risk of the test.

III. PROBABILITY MEASURE TRANSFORM: REVIEW

In this section, we review the principles of the probability measure transform [15]–[20] in the context of the considered binary hypothesis testing problem. We redefine the measure-transformed mean vector and covariance matrix and show their relation to higher-order statistical moments. Moreover, we reformulate their empirical estimators and restate the conditions for strong consistency and robustness to outliers. These quantities will be used in the following section to construct the measure-transformed Gaussian quasi likelihood ratio test.

A. Probability measure transform

Definition 1. Given a non-negative function $u : \mathbb{C}^p \rightarrow \mathbb{R}_+$ satisfying

$$0 < \mathbb{E}[u(\mathbf{X}); P_{\mathbf{X};H}] < \infty, \tag{5}$$

a transform on $P_{\mathbf{x};H}$ is defined via the relation:

$$Q_{\mathbf{x};H}^{(u)}(A) \triangleq T_u[P_{\mathbf{x};H}](A) = \int_A \varphi_u(\mathbf{x}; H) dP_{\mathbf{x};H}(\mathbf{x}), \quad (6)$$

where $A \in \mathcal{S}_{\mathcal{X}}$ and

$$\varphi_u(\mathbf{x}; H) \triangleq \frac{u(\mathbf{x})}{E[u(\mathbf{X}); P_{\mathbf{x};H}]}. \quad (7)$$

The function $u(\cdot)$ is called the MT-function.

Similarly to Proposition 1 in [17], the parameterized transformation (6) has the following properties:

Proposition 1 (Properties of the transform). *Let $Q_{\mathbf{x};H}^{(u)}$ be defined by relation (6). Then*

- 1) $Q_{\mathbf{x};H}^{(u)}$ is a probability measure on $\mathcal{S}_{\mathcal{X}}$.
- 2) $Q_{\mathbf{x};H}^{(u)}$ is absolutely continuous w.r.t. $P_{\mathbf{x};H}$, with Radon-Nikodym derivative [33]:

$$\frac{dQ_{\mathbf{x};H}^{(u)}(\mathbf{x})}{dP_{\mathbf{x};H}(\mathbf{x})} = \varphi_u(\mathbf{x}; H). \quad (8)$$

[Proof: see Proposition 1 in [17]]

The MT-function $u(\cdot)$ is the generating function of the probability measure $Q_{\mathbf{x};H}^{(u)}$.

B. The MT-mean and MT-covariance

According to (8) the mean vector and covariance matrix of \mathbf{X} under $Q_{\mathbf{x};H}^{(u)}$ are given by:

$$\boldsymbol{\mu}_{\mathbf{x};H}^{(u)} \triangleq E[\mathbf{X}\varphi_u(\mathbf{X}; H); P_{\mathbf{x};H}] \quad (9)$$

and

$$\boldsymbol{\Sigma}_{\mathbf{x};H}^{(u)} \triangleq E[\mathbf{X}\mathbf{X}^H\varphi_u(\mathbf{X}; H); P_{\mathbf{x};H}] - \boldsymbol{\mu}_{\mathbf{x};H}^{(u)}\boldsymbol{\mu}_{\mathbf{x};H}^{(u)H}, \quad (10)$$

respectively. Equations (9) and (10) imply that $\boldsymbol{\mu}_{\mathbf{x};H}^{(u)}$ and $\boldsymbol{\Sigma}_{\mathbf{x};H}^{(u)}$ are weighted mean and covariance of \mathbf{X} under $P_{\mathbf{x};H}$, with the weighting function $\varphi_u(\cdot; \cdot)$ defined in (7). By modifying the MT-function $u(\cdot)$, such that the condition (5) is satisfied, the MT-mean and MT-covariance under $Q_{\mathbf{x};H}^{(u)}$ are modified. In particular, by choosing $u(\cdot)$ to be any non-zero constant valued function we have $Q_{\mathbf{x};H}^{(u)} = P_{\mathbf{x};H}$, for which the standard mean vector $\boldsymbol{\mu}_{\mathbf{x};H}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{x};H}$ are obtained. Alternatively, when $u(\cdot)$ is a non-constant analytic function, which has a convergent Taylor series expansion, the resulting MT-mean and MT-covariance involve *higher-order statistical moments* of $P_{\mathbf{x};H}$.

C. The empirical MT-mean and MT-covariance

Given a sequence of N i.i.d. samples from $P_{\mathbf{X};H}$ the empirical estimators of $\boldsymbol{\mu}_{\mathbf{X};H}^{(u)}$ and $\boldsymbol{\Sigma}_{\mathbf{X};H}^{(u)}$ are defined as:

$$\hat{\boldsymbol{\mu}}_{\mathbf{X}}^{(u)} \triangleq \sum_{n=1}^N \mathbf{X}_n \hat{\varphi}_u(\mathbf{X}_n) \quad (11)$$

and

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{X}}^{(u)} \triangleq \sum_{n=1}^N \mathbf{X}_n \mathbf{X}_n^H \hat{\varphi}_u(\mathbf{X}_n) - \hat{\boldsymbol{\mu}}_{\mathbf{X}}^{(u)} \hat{\boldsymbol{\mu}}_{\mathbf{X}}^{(u)H}, \quad (12)$$

respectively, where

$$\hat{\varphi}_u(\mathbf{X}_n) \triangleq \frac{u(\mathbf{X}_n)}{\sum_{n=1}^N u(\mathbf{X}_n)}. \quad (13)$$

According to Proposition 2 in [17], if $E[\|\mathbf{X}\|^2 u(\mathbf{X}); P_{\mathbf{X};H}] < \infty$ then $\hat{\boldsymbol{\mu}}_{\mathbf{X}}^{(u)} \xrightarrow[N \rightarrow \infty]{\text{w.p. } 1} \boldsymbol{\mu}_{\mathbf{X};H}^{(u)}$ and $\hat{\boldsymbol{\Sigma}}_{\mathbf{X}}^{(u)} \xrightarrow[N \rightarrow \infty]{\text{w.p. } 1} \boldsymbol{\Sigma}_{\mathbf{X};H}^{(u)}$, where “ $\xrightarrow{\text{w.p. } 1}$ ” denotes convergence with probability (w.p.) 1 [36]. Note that for $u(\mathbf{X}) \equiv 1$ the estimators $\hat{\boldsymbol{\mu}}_{\mathbf{X}}^{(u)}$ and $\frac{N}{N-1} \hat{\boldsymbol{\Sigma}}_{\mathbf{X}}^{(u)}$ reduce to the standard unbiased sample mean vector (SMV) and sample covariance matrix (SCM), respectively.

D. Robustness to outliers

Robustness of the empirical MT-covariance (12) to outliers was studied in [17] using its influence function [37] which describes the effect on the estimator of an infinitesimal contamination at some point $\mathbf{y} \in \mathbb{C}^p$. An estimator is said to be B-robust if its influence function is bounded [37]. Similarly to the proof of Proposition 3 in [17] it can be shown that if there exists a finite positive constant $M \in \mathbb{R}$, such that for all $\mathbf{y} \in \mathbb{C}^p$:

$$u(\mathbf{y}) \leq M \quad \text{and} \quad u(\mathbf{y}) \|\mathbf{y}\|^2 \leq M, \quad (14)$$

then the influence functions of both (11) and (12) are bounded.

IV. THE MEASURE-TRANSFORMED GAUSSIAN QUASI LIKELIHOOD RATIO TEST

In this section, we extend the GQLRT (4) by applying the transformation (6) to the underlying probability measure $P_{\mathbf{X};H}$. Similarly to the standard GQLRT, given a sequence of samples from $P_{\mathbf{X};H}$, the proposed MT-GQLRT compares the empirical KLDs between the transformed distribution $Q_{\mathbf{X};H}^{(u)}$ and two complex circular Gaussian probability measures $\Phi_{\mathbf{X};H_k}^{(u)}$, $k = 0, 1$, that are characterized by the MT-mean vectors $\boldsymbol{\mu}_{\mathbf{X};H_k}^{(u)}$, $k = 0, 1$ and the MT-covariance matrices $\boldsymbol{\Sigma}_{\mathbf{X};H_k}^{(u)}$, $k = 0, 1$, which are assumed to be known (up to some redundant constants). Regularity conditions for asymptotic normality of the proposed test statistic are derived. When these conditions are satisfied we show that the resulting test is consistent and

derive its asymptotic size and power. Optimal selection of the MT-function $u(\cdot)$ out of some parametric class of functions is also discussed.

A. The MT-GQLRT

The KLD between $Q_{\mathbf{x};H}^{(u)}$ and $\Phi_{\mathbf{x};H_k}^{(u)}$, $k \in \{0, 1\}$ is defined as [9]:

$$D_{\text{KL}} \left[Q_{\mathbf{x};H}^{(u)} \parallel \Phi_{\mathbf{x};H_k}^{(u)} \right] \triangleq \mathbb{E} \left[\log \frac{q^{(u)}(\mathbf{X}; H)}{\phi^{(u)}(\mathbf{X}; H_k)}; Q_{\mathbf{x};H}^{(u)} \right], \quad (15)$$

where $q^{(u)}(\mathbf{x}; H)$ and

$$\phi^{(u)}(\mathbf{x}; H_k) \triangleq \det^{-1} \left[\pi \Sigma_{\mathbf{x};H_k}^{(u)} \right] \exp \left(- \left(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x};H_k}^{(u)} \right)^H \left(\Sigma_{\mathbf{x};H_k}^{(u)} \right)^{-1} \left(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x};H_k}^{(u)} \right) \right) \quad (16)$$

are the density functions of $Q_{\mathbf{x};H}^{(u)}$ and $\Phi_{\mathbf{x};H_k}^{(u)}$, respectively. By (15) and (16), one can verify that when $\phi^{(u)}(\cdot; H_0) \neq \phi^{(u)}(\cdot; H_1)$, the difference $D_{\text{KL}} \left[Q_{\mathbf{x};H}^{(u)} \parallel \Phi_{\mathbf{x};H_0}^{(u)} \right] - D_{\text{KL}} \left[Q_{\mathbf{x};H}^{(u)} \parallel \Phi_{\mathbf{x};H_1}^{(u)} \right]$ will be negative under H_0 and positive under H_1 . Hence, similarly to the standard GQLRT, this motivates the use of the empirical estimate of this difference as a test statistic for testing H_0 versus H_1 .

According to (8), the divergence $D_{\text{KL}} \left[Q_{\mathbf{x};H}^{(u)} \parallel \Phi_{\mathbf{x};H_k}^{(u)} \right]$, $k \in \{0, 1\}$ can be estimated using only samples from $P_{\mathbf{x};H}$. Therefore, similarly to (11) and (12), an empirical estimate of (15) given a sequence of samples $\mathbf{X}_n, n = 1, \dots, N$ from $P_{\mathbf{x};H}$, is defined as:

$$\hat{D}_{\text{KL}} \left[Q_{\mathbf{x};H}^{(u)} \parallel \Phi_{\mathbf{x};H_k}^{(u)} \right] \triangleq \sum_{n=1}^N \hat{\varphi}_u(\mathbf{X}_n) \log \frac{q^{(u)}(\mathbf{X}_n; H)}{\phi^{(u)}(\mathbf{X}_n; H_k)},$$

where $\hat{\varphi}_u(\cdot)$ is defined in (13). Thus, similarly to (3), the proposed test statistic, which is independent of the unknown transformed density function $q^{(u)}(\mathbf{x}; H)$, is defined as:

$$\begin{aligned} T_u &\triangleq \hat{D}_{\text{KL}} \left[Q_{\mathbf{x};H}^{(u)} \parallel \Phi_{\mathbf{x};H_0}^{(u)} \right] - \hat{D}_{\text{KL}} \left[Q_{\mathbf{x};H}^{(u)} \parallel \Phi_{\mathbf{x};H_1}^{(u)} \right] \\ &= \sum_{n=1}^N \hat{\varphi}_u(\mathbf{X}_n) \psi_u(\mathbf{X}_n) \\ &= \left(D_{\text{LD}} \left[\hat{\Sigma}_{\mathbf{x}}^{(u)} \parallel \Sigma_{\mathbf{x};H_0}^{(u)} \right] + \left\| \hat{\boldsymbol{\mu}}_{\mathbf{x}}^{(u)} - \boldsymbol{\mu}_{\mathbf{x};H_0}^{(u)} \right\|_{(\Sigma_{\mathbf{x};H_0}^{(u)})^{-1}}^2 \right) \\ &\quad - \left(D_{\text{LD}} \left[\hat{\Sigma}_{\mathbf{x}}^{(u)} \parallel \Sigma_{\mathbf{x};H_1}^{(u)} \right] + \left\| \hat{\boldsymbol{\mu}}_{\mathbf{x}}^{(u)} - \boldsymbol{\mu}_{\mathbf{x};H_1}^{(u)} \right\|_{(\Sigma_{\mathbf{x};H_1}^{(u)})^{-1}}^2 \right), \end{aligned} \quad (17)$$

where

$$\psi_u(\mathbf{X}) \triangleq \log \frac{\phi^{(u)}(\mathbf{X}; H_1)}{\phi^{(u)}(\mathbf{X}; H_0)}, \quad (18)$$

and the operators $D_{\text{LD}}[\cdot \parallel \cdot]$ and $\|\cdot\|_{(\cdot)}$ are defined below (3). The decision rule based on the test statistic (17) is:

$$T_u \underset{H_0}{\overset{H_1}{\gtrless}} t, \quad (19)$$

where $t \in \mathbb{R}$ denotes a threshold value. By modifying the MT-function $u(\cdot)$ such that condition (5) is satisfied the MT-GQLRT is modified, resulting in a family of tests generalizing the GQLRT described in Section II. In particular, if $u(\cdot)$ is any non-zero constant function over \mathcal{X} , then $Q_{\mathbf{X};H}^{(u)} = P_{\mathbf{X};H}$ and the standard non-robust GQLRT is obtained which only involves first and second-order statistical moments. Otherwise, when $u(\cdot)$ is non-constant analytic function that satisfies condition (14), the resulting test is outlier resilient and involves higher-order statistical moments.

B. Asymptotic performance analysis

Here, we study the asymptotic decision performance of the proposed MT-GQLRT (19). For simplicity, we assume that a sequence of i.i.d. samples \mathbf{X}_n , $n = 1, \dots, N$ from $P_{\mathbf{X};H}$ is available.

Theorem 1 (Asymptotic normality). *Assume that the following conditions are satisfied:*

A-1) $\mu_{\mathbf{X};H_0}^{(u)} \neq \mu_{\mathbf{X};H_1}^{(u)}$ or $\Sigma_{\mathbf{X};H_0}^{(u)} \neq \Sigma_{\mathbf{X};H_1}^{(u)}$.

A-2) $\Sigma_{\mathbf{X};H_0}^{(u)}$ and $\Sigma_{\mathbf{X};H_1}^{(u)}$ are non-singular.

A-3) $E[u^2(\mathbf{X}); P_{\mathbf{X};H}]$ and $E[\|\mathbf{X}\|^4 u^2(\mathbf{X}); P_{\mathbf{X};H}]$ are finite for $H = H_0$ and $H = H_1$.

Then,

$$\frac{T_u - \eta_H^{(u)}}{\sqrt{\lambda_H^{(u)}}} \xrightarrow[N \rightarrow \infty]{D} \mathcal{N}(0, 1) \quad \forall H \in \mathcal{H},$$

where “ \xrightarrow{D} ” denotes convergence in distribution [36],

$$\eta_H^{(u)} \triangleq E[\varphi_u(\mathbf{X}; H) \psi_u(\mathbf{X}); P_{\mathbf{X};H}] \quad (20)$$

and

$$\lambda_H^{(u)} \triangleq \frac{1}{N} E\left[\varphi_u^2(\mathbf{X}; H) \left(\psi_u(\mathbf{X}) - \eta_H^{(u)}\right)^2; P_{\mathbf{X};H}\right]. \quad (21)$$

[A proof is given in Appendix A]

Corollary 1 (Asymptotic size and power). *Assume that the conditions stated in Theorem 1 are satisfied.*

The asymptotic size and power of the decision rule (19) are given by:

$$\alpha_u \triangleq Q\left(\frac{t - \eta_{H_0}^{(u)}}{\sqrt{\lambda_{H_0}^{(u)}}}\right) \quad \text{and} \quad \beta_u \triangleq Q\left(\frac{t - \eta_{H_1}^{(u)}}{\sqrt{\lambda_{H_1}^{(u)}}}\right), \quad (22)$$

respectively, where $Q(\cdot)$ denotes the tail probability of the standard normal distribution [38].

Corollary 2 (Consistency). *Assume that the conditions in Theorem 1 are satisfied. Then, for any fixed asymptotic size the asymptotic power of the test (19) satisfies $\beta_u \rightarrow 1$ as $N \rightarrow \infty$.*

In the following Proposition, strongly consistent estimates of the asymptotic size and power (22) are constructed based on two i.i.d. training sequences from $P_{\mathbf{X};H_0}$ and $P_{\mathbf{X};H_1}$. These quantities will be used in the sequel for optimal selection of the MT-function.

Proposition 2 (Empirical asymptotic size and power). *Let $\mathbf{X}_n^{(k)}$, $n = 1, \dots, N_k$, $k = 0, 1$ denote sequences of i.i.d. samples from $P_{\mathbf{X};H_0}$ and $P_{\mathbf{X};H_1}$, respectively. Define the empirical asymptotic size and power:*

$$\hat{\alpha}_u \triangleq Q \left(\frac{t - \hat{\eta}_{H_0}^{(u)}}{\sqrt{\hat{\lambda}_{H_0}^{(u)}}} \right) \quad \text{and} \quad \hat{\beta}_u \triangleq Q \left(\frac{t - \hat{\eta}_{H_1}^{(u)}}{\sqrt{\hat{\lambda}_{H_1}^{(u)}}} \right), \quad (23)$$

respectively, where

$$\hat{\eta}_{H_k}^{(u)} \triangleq \sum_{n=1}^{N_k} \hat{\varphi}_u \left(\mathbf{X}_n^{(k)} \right) \psi_u \left(\mathbf{X}_n^{(k)} \right) \quad (24)$$

and

$$\hat{\lambda}_{H_k}^{(u)} \triangleq \frac{N_k}{N} \sum_{n=1}^{N_k} \hat{\varphi}_u^2 \left(\mathbf{X}_n^{(k)} \right) \left(\psi_u \left(\mathbf{X}_n^{(k)} \right) - \hat{\eta}_{H_k}^{(u)} \right)^2. \quad (25)$$

Assume that condition A-1 - A-3 stated in Theorem 1 are satisfied. Then,

$$\hat{\alpha}_u \xrightarrow[N_0 \rightarrow \infty]{w.p.1} \alpha_u \quad \text{and} \quad \hat{\beta}_u \xrightarrow[N_1 \rightarrow \infty]{w.p.1} \beta_u.$$

[A proof is given in Appendix C]

C. Selection of the MT-function

We propose to specify the MT-function within some parametric family $\{u(\mathbf{X}; \boldsymbol{\omega}), \boldsymbol{\omega} \in \boldsymbol{\Omega} \subseteq \mathbb{C}^r\}$ that satisfies the conditions stated in Definition 1 and Theorem 1. For example, in order to gain resilience against outliers, the Gaussian family of functions that satisfy condition (14) is a natural choice. An optimal choice of the MT-function parameter $\boldsymbol{\omega}$ would be the one that maximizes the empirical asymptotic power in (23) at a fixed empirical asymptotic size $\hat{\alpha}_u = \alpha$, i.e., we maximize the following objective function:

$$\hat{\beta}_u^{(\alpha)}(\boldsymbol{\omega}) = Q \left(\frac{\hat{\eta}_{H_0}^{(u)}(\boldsymbol{\omega}) - \hat{\eta}_{H_1}^{(u)}(\boldsymbol{\omega}) + \sqrt{\hat{\lambda}_{H_0}^{(u)}(\boldsymbol{\omega})} Q^{-1}(\alpha)}{\sqrt{\hat{\lambda}_{H_1}^{(u)}(\boldsymbol{\omega})}} \right). \quad (26)$$

V. BAYESIAN EXTENSION OF THE MT-GQLRT

In this section, we develop a Bayesian extension of the proposed MT-GQLRT (19). We consider the measure space $(\mathcal{X}, \mathcal{S}_{\mathcal{X}}, P_{\mathbf{X}|H})$, where $\mathcal{X} \subseteq \mathbb{C}^p$ is the observation space of a random vector \mathbf{X} , $\mathcal{S}_{\mathcal{X}}$ is a σ -algebra over \mathcal{X} and $P_{\mathbf{X}|H}$ is a probability measure on $\mathcal{S}_{\mathcal{X}}$ conditioned on a random symbolic variable H that takes values in the pair set $\mathcal{H} \triangleq \{H_0, H_1\}$ with known a-priori probabilities, P_{H_0} and P_{H_1} , respectively.

Given a sequence of samples from the conditional distribution $P_{\mathbf{x}|H}$ we consider the problem of testing between the null and alternative hypotheses $H = H_0$ and $H = H_1$, respectively, when $P_{\mathbf{x}|H_0}$ and $P_{\mathbf{x}|H_1}$ are *unknown*. Here, the measure transformation (6) is applied to the conditional measure $P_{\mathbf{x}|H}$ instead of the unconditional measure $P_{\mathbf{x};H}$. The transformed conditional probability measure and its corresponding measure-transformed mean and covariance will be denoted by $Q_{\mathbf{x}|H}^{(u)}$, $\boldsymbol{\mu}_{\mathbf{x}|H}^{(u)}$ and $\boldsymbol{\Sigma}_{\mathbf{x}|H}^{(u)}$, respectively. Note that all the properties stated in Section III for the unconditional case apply here by simply replacing $P_{\mathbf{x};H}$ with the conditional measure $P_{\mathbf{x}|H}$.

A. The Bayesian MT-GQLRT

Let $\Phi_{\mathbf{x}|H}^{(u)}$ denote a complex circular Gaussian probability distribution [34] characterized by the conditional MT-mean vector $\boldsymbol{\mu}_{\mathbf{x}|H}^{(u)}$ and the conditional MT-covariance matrix $\boldsymbol{\Sigma}_{\mathbf{x}|H}^{(u)}$. Similarly to (17), given a sequence of samples $\mathbf{X}_n, n = 1, \dots, N$ from $P_{\mathbf{x}|H}$, the proposed test statistic compares the empirical KLDs between $Q_{\mathbf{x}|H}^{(u)}$ and $\Phi_{\mathbf{x}|H_k}^{(u)}$, $k = 0, 1$ and takes the form:

$$\begin{aligned} S_u &\triangleq \hat{D}_{\text{KL}} \left[Q_{\mathbf{x}|H}^{(u)} \parallel \Phi_{\mathbf{x}|H_0}^{(u)} \right] - \hat{D}_{\text{KL}} \left[Q_{\mathbf{x}|H}^{(u)} \parallel \Phi_{\mathbf{x}|H_1}^{(u)} \right] \\ &= \sum_{n=1}^N \hat{\varphi}_u(\mathbf{X}_n) \xi_u(\mathbf{X}_n) \\ &= \left(D_{\text{LD}} \left[\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u)} \parallel \boldsymbol{\Sigma}_{\mathbf{x}|H_0}^{(u)} \right] + \left\| \hat{\boldsymbol{\mu}}_{\mathbf{x}}^{(u)} - \boldsymbol{\mu}_{\mathbf{x}|H_0}^{(u)} \right\|_{(\boldsymbol{\Sigma}_{\mathbf{x}|H_0}^{(u)})^{-1}}^2 \right) \\ &\quad - \left(D_{\text{LD}} \left[\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u)} \parallel \boldsymbol{\Sigma}_{\mathbf{x}|H_1}^{(u)} \right] + \left\| \hat{\boldsymbol{\mu}}_{\mathbf{x}}^{(u)} - \boldsymbol{\mu}_{\mathbf{x}|H_1}^{(u)} \right\|_{(\boldsymbol{\Sigma}_{\mathbf{x}|H_1}^{(u)})^{-1}}^2 \right), \end{aligned} \quad (27)$$

where

$$\xi_u(\mathbf{X}) \triangleq \log \frac{\phi^{(u)}(\mathbf{X}_n|H_1)}{\phi^{(u)}(\mathbf{X}_n|H_0)},$$

$\phi^{(u)}(\cdot|H)$ is the density function of $\Phi_{\mathbf{x}|H}^{(u)}$ and the operators $D_{\text{LD}}[\cdot|\cdot]$ and $\|\cdot\|_{(\cdot)}$ are defined below (3).

The decision rule based on the test statistic (27) is:

$$S_u \underset{H_0}{\overset{H_1}{\gtrless}} t, \quad (28)$$

where $t \in \mathbb{R}$ denotes a threshold value.

B. Asymptotic performance analysis

Here, we study the asymptotic performance of the Bayesian MT-GQLRT (28). As in the non-Bayesian case, we assume that a sequence of i.i.d. samples $\mathbf{X}_n, n = 1, \dots, N$ from the conditional distribution $P_{\mathbf{x}|H}$ is available. Straight forward extension of Theorem 1 (asymptotic normality of the test statistic) to

the considered Bayesian case can be obtained here by replacing $P_{\mathbf{X};H}$, $\boldsymbol{\mu}_{\mathbf{X};H}^{(u)}$ and $\boldsymbol{\Sigma}_{\mathbf{X};H}^{(u)}$ with $P_{\mathbf{X}|H}$, $\boldsymbol{\mu}_{\mathbf{X}|H}^{(u)}$ and $\boldsymbol{\Sigma}_{\mathbf{X}|H}^{(u)}$, respectively. Under this extension, the asymptotic Bayes risk and its empirical estimate are stated in the following propositions.

Proposition 3 (Asymptotic Bayes risk). *Assume that the following conditions are satisfied:*

B-1) $\boldsymbol{\mu}_{\mathbf{X}|H_0}^{(u)} \neq \boldsymbol{\mu}_{\mathbf{X}|H_1}^{(u)}$ or $\boldsymbol{\Sigma}_{\mathbf{X}|H_0}^{(u)} \neq \boldsymbol{\Sigma}_{\mathbf{X}|H_1}^{(u)}$.

B-2) $\boldsymbol{\Sigma}_{\mathbf{X}|H_0}^{(u)}$ and $\boldsymbol{\Sigma}_{\mathbf{X}|H_1}^{(u)}$ are non-singular.

B-3) $\mathbb{E} [u^2(\mathbf{X}); P_{\mathbf{X}|H}]$ and $\mathbb{E} [\|\mathbf{X}\|^4 u^2(\mathbf{X}); P_{\mathbf{X}|H}]$ are finite for $H = H_0$ and $H = H_1$.

For a loss function,

$$L(H, \hat{H}) = \begin{cases} L_{10}, & \text{if } \hat{H} = H_1 \text{ and } H = H_0 \\ L_{01}, & \text{if } \hat{H} = H_0 \text{ and } H = H_1, \\ 0, & \text{otherwise} \end{cases}$$

where \hat{H} denotes the outcome of the test (28), the asymptotic Bayes risk can be written as:

$$R^{(u)}(t) \triangleq L_{10} P_{H_0} Q \left(\frac{t - \kappa_{H_0}^{(u)}}{\sqrt{\gamma_{H_0}^{(u)}}} \right) + L_{01} P_{H_1} Q \left(\frac{\kappa_{H_1}^{(u)} - t}{\sqrt{\gamma_{H_1}^{(u)}}} \right), \quad (29)$$

where,

$$\kappa_{H_k}^{(u)} \triangleq \mathbb{E} [\varphi_u(\mathbf{X}|H_k) \xi_u(\mathbf{X}); P_{\mathbf{X}|H_k}], \quad k = 0, 1,$$

$$\gamma_{H_k}^{(u)} \triangleq \frac{1}{N} \mathbb{E} \left[\varphi_u^2(\mathbf{X}|H_k) \left(\xi_u(\mathbf{X}) - \kappa_{H_k}^{(u)} \right)^2; P_{\mathbf{X}|H_k} \right], \quad k = 0, 1$$

and $\varphi_u(\mathbf{X}|H_k) \triangleq \frac{dQ_{\mathbf{X}|H}^{(u)}(\mathbf{x})}{dP_{\mathbf{X}|H}(\mathbf{x})} = u(\mathbf{X}) / \mathbb{E}[u(\mathbf{X}); P_{\mathbf{X}|H}]$ is the Radon-Nikodym derivative of $Q_{\mathbf{X}|H}^{(u)}$ w.r.t. $P_{\mathbf{X}|H}$.

In the following Proposition, a strongly consistent estimate of the asymptotic Bayes risk (29) is constructed based on two i.i.d. sequences from the conditional distributions $P_{\mathbf{X}|H_0}$ and $P_{\mathbf{X}|H_1}$. This quantity will be used in the sequel for optimal selection of the MT-function.

Proposition 4 (Empirical asymptotic Bayes risk). *Let $\mathbf{X}_n^{(k)}$, $n = 1, \dots, N_k$, $k = 0, 1$ denote sequences of i.i.d. samples from $P_{\mathbf{X}|H_0}$ and $P_{\mathbf{X}|H_1}$, respectively. Define the empirical asymptotic Bayes risk:*

$$\hat{R}^{(u)}(t) \triangleq L_{10} P_{H_0} Q \left(\frac{t - \hat{\kappa}_{H_0}^{(u)}}{\sqrt{\hat{\gamma}_{H_0}^{(u)}}} \right) + L_{01} P_{H_1} Q \left(\frac{\hat{\kappa}_{H_1}^{(u)} - t}{\sqrt{\hat{\gamma}_{H_1}^{(u)}}} \right), \quad (30)$$

where

$$\hat{\kappa}_{H_k}^{(u)} \triangleq \sum_{n=1}^{N_k} \hat{\varphi}_u(\mathbf{X}_n^{(k)}) \xi_u(\mathbf{X}_n^{(k)}),$$

$$\hat{\gamma}_{H_k}^{(u)} \triangleq \frac{N_k}{N} \sum_{n=1}^{N_k} \hat{\varphi}_u^2(\mathbf{X}_n^{(k)}) \left(\xi_u(\mathbf{X}_n^{(k)}) - \hat{\kappa}_{H_k}^{(u)} \right)^2$$

and $\hat{\varphi}_u(\cdot)$ is defined as in (13). Assume that conditions B-1 - B-3 stated in Proposition 3 is satisfied.

Then, $\hat{R}^{(u)} \xrightarrow[N_0, N_1 \rightarrow \infty]{w.p.1} R^{(u)}$.

[A proof is given in Appendix D]

In the following proposition, a necessary and sufficient condition for existence and uniqueness of an optimal threshold minimizing the empirical asymptotic Bayes risk (30) is derived. A closed form expression of this threshold is also presented.

Proposition 5 (Optimal threshold). Assume that $\hat{\gamma}_{H_0}^{(u)} \neq \hat{\gamma}_{H_1}^{(u)}$ ¹. Define

$$\hat{s}^{(u)} \triangleq \left(\hat{\kappa}_{H_0}^{(u)} - \hat{\kappa}_{H_1}^{(u)} \right)^2 - 2 \left(\hat{\gamma}_{H_0}^{(u)} - \hat{\gamma}_{H_1}^{(u)} \right) \log \frac{L_{10} P_{H_0} \sqrt{\hat{\gamma}_{H_1}^{(u)}}}{L_{01} P_{H_1} \sqrt{\hat{\gamma}_{H_0}^{(u)}}}.$$

A global minimum of the empirical asymptotic Bayes risk (30) exists and given by

$$t_{opt}^{(u)} \triangleq \frac{\hat{\gamma}_{H_0}^{(u)} \hat{\kappa}_{H_1}^{(u)} - \hat{\gamma}_{H_1}^{(u)} \hat{\kappa}_{H_0}^{(u)} - \sqrt{\hat{\gamma}_{H_0}^{(u)} \hat{\gamma}_{H_1}^{(u)} \hat{s}^{(u)}}}{\hat{\gamma}_{H_0}^{(u)} - \hat{\gamma}_{H_1}^{(u)}}, \quad (31)$$

if and only if

C-1) $\hat{s}^{(u)} \geq 0$.

and

C-2) the empirical Bayes risk (30) satisfies $\hat{R}^{(u)}(t_{opt}^{(u)}) < \min(L_{10} P_{H_0}, L_{01} P_{H_1})$.

[A proof is given in Appendix E]

C. Optimal selection of the MT-function

Similarly to the non-Bayesian case, we propose to specify the MT-function within some parametric family $\{u(\mathbf{X}; \omega), \omega \in \Omega \subseteq \mathbb{C}^r\}$ of functions that have strictly positive and finite expectation w.r.t. the conditional distribution of the data and satisfy conditions B-1 - B-3 stated in Proposition 3. An optimal choice of the MT-function parameter ω minimizes the empirical asymptotic Bayes risk (30) evaluated at the optimal threshold (31).

VI. EXAMPLES

In this section we illustrate the proposed MT-GQLRT and its Bayesian extension to random signal detection and deterministic signal classification, respectively. Other applications of these tests for Bayesian and non-Bayesian random signal classification are detailed in the conference papers [39], [40].

¹Notice that when \mathbf{X} is a continuous random vector, this assumption satisfied almost surely [33].

A. non-Bayesian MT-GQLRT: Signal detection

We consider the following signal detection problem:

$$\begin{aligned} H_0 &: \mathbf{X}_n = \mathbf{W}_n, \quad n = 1, \dots, N, \\ H_1 &: \mathbf{X}_n = S_n \mathbf{a} + \mathbf{W}_n, \quad n = 1, \dots, N, \end{aligned} \quad (32)$$

where n is a discrete time index, $\mathbf{X}_n \in \mathbb{C}^p$ is an observation vector with $p > 1$ components, $S_n \in \mathbb{C}$ is a first-order stationary zero-mean random signal, and $\mathbf{a} \in \mathbb{C}^p$ is a known unit norm deterministic vector. The vector $\mathbf{W}_n \in \mathbb{C}^p$ is a first-order stationary additive noise that is statistically independent of S_n .

In order to derive the MT-GQLRT for this decision problem we specify the MT-function in the set:

$$\left\{ u(\mathbf{x}) = v\left(\mathbf{P}_\mathbf{a}^\perp \mathbf{x}\right), \quad v: \mathbb{C}^p \rightarrow \mathbb{R}_+ \right\}, \quad (33)$$

where $\mathbf{P}_\mathbf{a}^\perp \triangleq \mathbf{I}_p - \mathbf{a}\mathbf{a}^H$ is the projection matrix into the subspace orthogonal to \mathbf{a} , and \mathbf{I}_p is a unit matrix of size of $p \times p$. Assuming that condition (5) is satisfied, one can verify using (7), (9), (10), (32) and (33) that the MT-mean and MT-covariance under the transformed probability measure $Q_{\mathbf{x};H_k}^{(u)}$, $k \in \{0, 1\}$ satisfy the following properties:

$$\boldsymbol{\mu}_{\mathbf{x};H_k}^{(u)} = \boldsymbol{\mu}_\mathbf{w}^{(u)}, \quad k = 0, 1 \quad (34)$$

and

$$\boldsymbol{\Sigma}_{\mathbf{x};H_0}^{(u)} = \boldsymbol{\Sigma}_\mathbf{w}^{(u)}, \quad \boldsymbol{\Sigma}_{\mathbf{x};H_1}^{(u)} = \sigma_S^2 \mathbf{a}\mathbf{a}^H + \boldsymbol{\Sigma}_\mathbf{w}^{(u)}, \quad (35)$$

where $\sigma_S^2 \triangleq \mathbb{E}[|S_n|^2; P_S]$ is the signal variance, and $\boldsymbol{\mu}_\mathbf{w}^{(u)}$ and $\boldsymbol{\Sigma}_\mathbf{w}^{(u)}$ are the MT-mean and MT-covariance of the noise component. Hence, by substituting (34) and (35) into (17) the resulting test statistic after subtraction of the observation-independent constant $c_1^{(u)} \triangleq \log \det(\boldsymbol{\Sigma}_{\mathbf{x};H_0}^{(u)} \boldsymbol{\Sigma}_{\mathbf{x};H_1}^{(u)-1})$ followed by normalization by the positive observation-independent constant $c_2^{(u)} \triangleq \sigma_S^2 / (1 + \sigma_S^2 \mathbf{a}^H \boldsymbol{\Sigma}_\mathbf{w}^{(u)-1} \mathbf{a})$ is given by:

$$T'_u \triangleq \frac{T_u - c_1^{(u)}}{c_2^{(u)}} = \mathbf{a}^H \left(\boldsymbol{\Sigma}_\mathbf{w}^{(u)} \right)^{-1} \hat{\mathbf{C}}_\mathbf{x}^{(u)} \left(\boldsymbol{\Sigma}_\mathbf{w}^{(u)} \right)^{-1} \mathbf{a}, \quad (36)$$

where $\hat{\mathbf{C}}_\mathbf{x}^{(u)} \triangleq \hat{\boldsymbol{\Sigma}}_\mathbf{x}^{(u)} + \hat{\boldsymbol{\mu}}_\mathbf{x}^{(u)} \hat{\boldsymbol{\mu}}_\mathbf{x}^{(u)H} - \boldsymbol{\mu}_\mathbf{w}^{(u)} \boldsymbol{\mu}_\mathbf{w}^{(u)H}$.

We further assume that the noise component has a density that is spherically contoured with stochastic representation [41]:

$$\mathbf{W}_n = \nu_n \mathbf{Z}_n, \quad (37)$$

where $\nu_n \in \mathbb{R}_{++}$ is a first-order stationary process and $\mathbf{Z}_n \in \mathbb{C}^p$ is a proper-complex wide-sense stationary Gaussian process with zero-mean and scaled unit covariance $\sigma_Z^2 \mathbf{I}$. The processes ν_n and \mathbf{Z}_n are assumed

to be statistically independent. Additionally, in order to mitigate the effect of outliers and involve higher-order statistical moments, we specify the MT-function in a subset of (33) that is comprised of zero-centred Gaussian functions parametrized by a width parameter ω , i.e.,

$$u_G(\mathbf{x}; \omega) = \exp\left(-\left\|\mathbf{P}_a^\perp \mathbf{x}\right\|^2 / \omega^2\right), \quad \omega \in \mathbb{R}_{++}. \quad (38)$$

Under the settings in (37) and (38) it can be shown using (9) and (10) that

$$\boldsymbol{\mu}_{\mathbf{w}}^{(u_G)}(\omega) = \mathbf{0} \quad (39)$$

and

$$\boldsymbol{\Sigma}_{\mathbf{w}}^{(u_G)}(\omega) = r_0(\omega) \mathbf{a} \mathbf{a}^H + r_1(\omega) \mathbf{I}, \quad (40)$$

respectively, where $r_0(\omega)$ and $r_1(\omega)$ are some strictly positive functions of ω . Hence, by substituting (39) and (40) into (36) followed by normalization by the observation-independent factor $c_3(\omega) \triangleq \frac{1}{(r_0(\omega) + r_1(\omega))^2}$, the MT-GQLRT (19) simplifies to:

$$T_{u_G}'' \triangleq T_{u_G}' / c_3(\omega) = \mathbf{a}^H \hat{\mathbf{C}}_{\mathbf{x}}^{(u_G)}(\omega) \mathbf{a} \underset{H_0}{\overset{H_1}{\gtrless}} t'', \quad (41)$$

where $\hat{\mathbf{C}}_{\mathbf{x}}^{(u_G)}(\omega) \triangleq \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{(u_G)}(\omega) + \hat{\boldsymbol{\mu}}_{\mathbf{x}}^{(u_G)}(\omega) \hat{\boldsymbol{\mu}}_{\mathbf{x}}^{(u_G)H}(\omega)$ and $t'' \triangleq (t - c_1^{(u_G)}) / (c_2^{(u_G)} c_3(\omega))$. Notice that when the vector \mathbf{a} represents a steering vector of a sensor array [42], the test statistic in (41) is a measure transformed version of Bartlett's beamformer [42].

Under the considered settings, it can be shown that the conditions stated in Theorem 1 are satisfied. The resulting asymptotic power (22) at a given asymptotic size $\alpha_{u_G} = \alpha$ takes the form:

$$\beta_{u_G}^{(\alpha)} = Q\left(\frac{\sqrt{\frac{1}{N} G_1(\omega)} Q^{-1}(\alpha) - G_2(\omega)}{\sqrt{\frac{1}{N} G_3(\omega)}}\right), \quad (42)$$

where

$$G_1(\omega) \triangleq \mathbb{E}\left[g\left(\omega, \sqrt{2}\tilde{\nu}\right)\left(2\tilde{\nu}^4 - 2L_1(\omega)\tilde{\nu}^2 + L_1^2(\omega)\right); P_\nu\right], \quad G_2(\omega) \triangleq \mathbb{E}\left[g\left(\omega, \tilde{\nu}\right)|S|^2; P_{\nu;S}\right],$$

$$G_3(\omega) \triangleq \mathbb{E}\left[g\left(\omega, \sqrt{2}\tilde{\nu}\right)\left(2\tilde{\nu}^4 + 4|S|^2\tilde{\nu}^2 + |S|^4 - 2L_2(\omega)\left(\tilde{\nu}^2 + |S|^2\right) + L_2^2(\omega)\right); P_{\nu;S}\right],$$

$$L_1(\omega) \triangleq \frac{\mathbb{E}\left[\tilde{\nu}^2 g\left(\omega, \tilde{\nu}\right); P_\nu\right]}{\mathbb{E}\left[g\left(\omega, \tilde{\nu}\right); P_\nu\right]}, \quad L_2(\omega) \triangleq L_1(\omega) + \mathbb{E}\left[|S|^2; P_S\right],$$

$g(\omega, \nu) \triangleq \left(\frac{\omega^2}{\omega^2 + \nu^2}\right)^{p-1}$ and $\tilde{\nu} \triangleq \sigma_Z^2 \nu$. Furthermore, its empirical estimate (26) is given by

$$\hat{\beta}_{u_G}^{(\alpha)}(\omega) = Q\left(\frac{\hat{\eta}_{H_0}^{(u_G)}(\omega) - \hat{\eta}_{H_1}^{(u_G)}(\omega) + \sqrt{\tilde{\lambda}_{H_0}^{(u_G)}(\omega)} Q^{-1}(\alpha)}{\sqrt{\tilde{\lambda}_{H_1}^{(u_G)}(\omega)}}\right), \quad (43)$$

where

$$\begin{aligned}\tilde{\eta}_{H_k}^{(u_G)}(\omega) &\triangleq \frac{\hat{\eta}_{H_k}^{(u_G)}(\omega) - c_1^{(u_G)}}{c_2^{(u_G)} c_3(\omega)} = \sum_{n=1}^{N_k} \hat{\varphi}_{u_G}(\mathbf{X}_n^{(k)}; \omega) \left| \mathbf{a}^H \mathbf{X}_n^{(k)} \right|^2, \\ \tilde{\lambda}_{H_k}^{(u_G)}(\omega) &\triangleq \frac{\hat{\lambda}_{H_k}^{(u_G)}(\omega)}{\left(c_2^{(u_G)} c_3(\omega) \right)^2} = \frac{N_k}{N} \sum_{n=1}^{N_k} \hat{\varphi}_{u_G}^2(\mathbf{X}_n^{(k)}; \omega) \left(\mathbf{a}^H \mathbf{X}_n^{(k)} \mathbf{X}_n^{(k)H} \mathbf{a} - \tilde{\eta}_{H_k}^{(u_G)}(\omega) \right)^2,\end{aligned}$$

$c_1^{(u_G)}$ and $c_2^{(u_G)}$ are defined above (36) and $c_3(\omega)$ is defined above (41).

Next, we study the robustness of the MT-GQLRT (41) to outliers. Notice that the Gaussian MT-function (38) does not decay outliers in the direction of the vector \mathbf{a} and does not satisfy the B-robustness condition (14) when the observations are proportional to \mathbf{a} . However, this condition is satisfied over a sufficiently large subset of \mathbb{C}^p , guaranteeing robustness of the empirical MT-mean and MT-covariance with sufficiently high probability. To see this, define the set $\mathcal{B}_\epsilon \triangleq \left\{ \mathbf{y} \in \mathbb{C}^p : \frac{|\mathbf{a}^H \mathbf{y}|^2}{\|\mathbf{a}\|^2 \|\mathbf{y}\|^2} \leq 1 - \epsilon \right\}$ where $\epsilon > 0$ is some fixed small positive constant. Clearly, $u_G(\mathbf{y}; \omega) \leq \exp\left(-\frac{\epsilon \|\mathbf{y}\|^2}{\omega^2}\right)$ for any $\mathbf{y} \in \mathcal{B}_\epsilon$ and for any fixed ω . Therefore, since $\exp\left(-\frac{\epsilon \|\mathbf{y}\|^2}{\omega^2}\right)$ and $\|\mathbf{y}\|^2 \exp\left(-\frac{\epsilon \|\mathbf{y}\|^2}{\omega^2}\right)$ are bounded over \mathbb{C}^p , the MT-function (38) must satisfy condition (14) over \mathcal{B}_ϵ . Finally, since the probability measure of \mathcal{B}_ϵ satisfies $P_{\mathbf{X};H}(\mathcal{B}_\epsilon) \approx 1$ for sufficiently small ϵ we conclude that the empirical MT-mean and MT-covariance, comprising (41), are robust to outliers with sufficiently high probability.

In the following simulation examples we evaluate the detection performance of the MT-GQLRT as compared to the omniscient LRT, the standard GQLRT (4), other robust GQLRT extensions, discussed in the following paragraph, and to the NSDD-GLRT [43]. The NSDD-GLRT is a robust generalized likelihood ratio test (GLRT) detector, which assumes that the signal samples in (32) are deterministic unknown and that the noise samples are zero-mean normally distributed with unknown variances.

Under the considered detection problem (32) one can verify using (3) that the test-statistic of the GQLRT reduces to $T_{GQLRT} = \mathbf{a}^H \hat{\mathbf{C}}_{\mathbf{X}} \mathbf{a}$, where $\hat{\mathbf{C}}_{\mathbf{X}} \triangleq \sum_{n=1}^N \mathbf{X}_n \mathbf{X}_n^H$ is the non-robust sample correlation matrix. Hence, other robust alternatives to the GQLRT can be obtained by replacing the non-robust sample correlation matrix with robust scatter estimates that assume zero location parameter, namely, the Tyler's scatter M-estimator [44] and the empirical sign covariance [45]. These GQLRT robust extensions will be called here Tyler-GQLRT and SGN-GQLRT, respectively. The maximum number of iterations and the stopping criterion of the Tyler's scatter M-estimator were set to 100 and $\|\hat{\mathbf{C}}_l^{(\text{Tyler})} - \hat{\mathbf{C}}_{l-1}^{(\text{Tyler})}\|_F / \|\hat{\mathbf{C}}_{l-1}^{(\text{Tyler})}\|_F < 10^{-6}$, respectively, where $\hat{\mathbf{C}}_l^{(\text{Tyler})}$ denotes the scatter M-estimation at iteration index l and $\|\cdot\|_F$ denotes the Frobenius norm [46]. Another, somewhat primitive, robust extension of the GQLRT is tested here that operates by applying GQLRT after passing the data through a zero-memory non-linear (ZMNL) function that suppresses outliers by clipping the amplitude

of the observations. This GQLRT extension is called here ZMNL-GQLRT. We use the same ZMNL preprocessing approach that has been applied to robustify the MUSIC algorithm [42] in [47]–[49].

In all simulation examples, the signal S_n in (32) is considered to be a BPSK signal with power σ_S^2 . The vector $\mathbf{a} \triangleq \frac{1}{\sqrt{p}} [1, e^{-i\pi \sin(\vartheta)}, \dots, e^{-i\pi(p-1) \sin(\vartheta)}]^T$ represents a steering vector of a uniform linear array with $p = 8$ elements with half wavelength spacing for a near-field narrow band signal arriving from DOA $\vartheta = \pi/3$ [Rad]. We considered two types of noise distributions with zero location parameter and isotropic dispersion $\sigma_Z^2 \mathbf{I}_p$: 1) Gaussian and 2) ϵ -contaminated Gaussian noise model [41] under which the texture component ν in (37) is a binary random variable satisfying $\nu = 1$ w.p. $1 - \epsilon$ and $\nu = \delta$ w.p. ϵ . The parameters ϵ and δ that control the heaviness of the noise tails were set to 0.25 and 10, respectively.

For each noise type we performed two simulations. In the first one, we compared the asymptotic power (42) to its empirical estimate (43) as a function of ω for sample size of $N = 300$. The empirical asymptotic power (43) was obtained using two i.i.d. training sequences from $P_{\mathbf{X};H_0}$ and $P_{\mathbf{X};H_1}$ containing $N_0 = N_1 = 3 \times 10^4$ samples. The signal-to-noise-ratio (SNR), defined here as $\text{SNR} \triangleq 10 \log_{10} \sigma_S^2 / \sigma_Z^2$ was set to -8 [dB] and -6 [dB] for the Gaussian and ϵ -contaminated Gaussian noise, respectively. Observing Figs. 1(a) and 2(a), one sees that due to the consistency of (43) the compared quantities are very close. This illustrates the reliability of the empirical asymptotic power for optimal choice of the of the MT-function parameter, as discussed in subsection IV-C.

In the second simulation, we compared the empirical, asymptotic (42) and empirical asymptotic (43) power of the MT-GQLRT to the empirical power of the GQLRT, Tyler-GQLRT, SGN-GQLRT, ZMNL-GQLRT, NSDD-GLRT and the omniscient LRT. The optimal Gaussian MT-function parameter ω_{opt} was obtained by minimizing (43) over $\Omega = [1, 100]$. The empirical power curves were obtained using 10^5 Monte-Carlo simulations. The SNR, sample size and the test size are used to index the performances as depicted in Figs. 1(b) - 1(d) for the Gaussian noise and in Figs. 2(b) - 2(d) for the ϵ -contaminated Gaussian noise. The power versus SNR was evaluated for a fixed test size equal to 0.05 and $N = 300$ i.i.d observations. The power versus sample size was evaluated for a fixed test size equal to 0.05 and $\text{SNR} = -12$ [dB] and $\text{SNR} = -11$ [dB] for the Gaussian and ϵ -contaminated Gaussian noise, respectively. The power versus test size (ROC curves) was evaluated for $N = 300$ i.i.d observations and $\text{SNR} = -8$ [dB]. Observing Figs. 1(b) - 1(d), one can notice that when the noise is Gaussian, the MT-GQLRT and the GQLRT obtain similar performance and outperform the other robust alternatives. This result is an outcome of the fact that the MT-GQLRT approaches the GQLRT as the width parameter of the Gaussian MT-function (38) approaches infinity. Observing Figs. 2(b) - 2(d), one sees that for the ϵ -contaminated Gaussian noise, the MT-GQLRT outperforms the non-robust GQLRT and the robust alternatives and

attains detection performance that are significantly closer to those obtained by the LRT that, unlike the MT-GQLRT, requires complete knowledge of the likelihood function under each hypothesis.

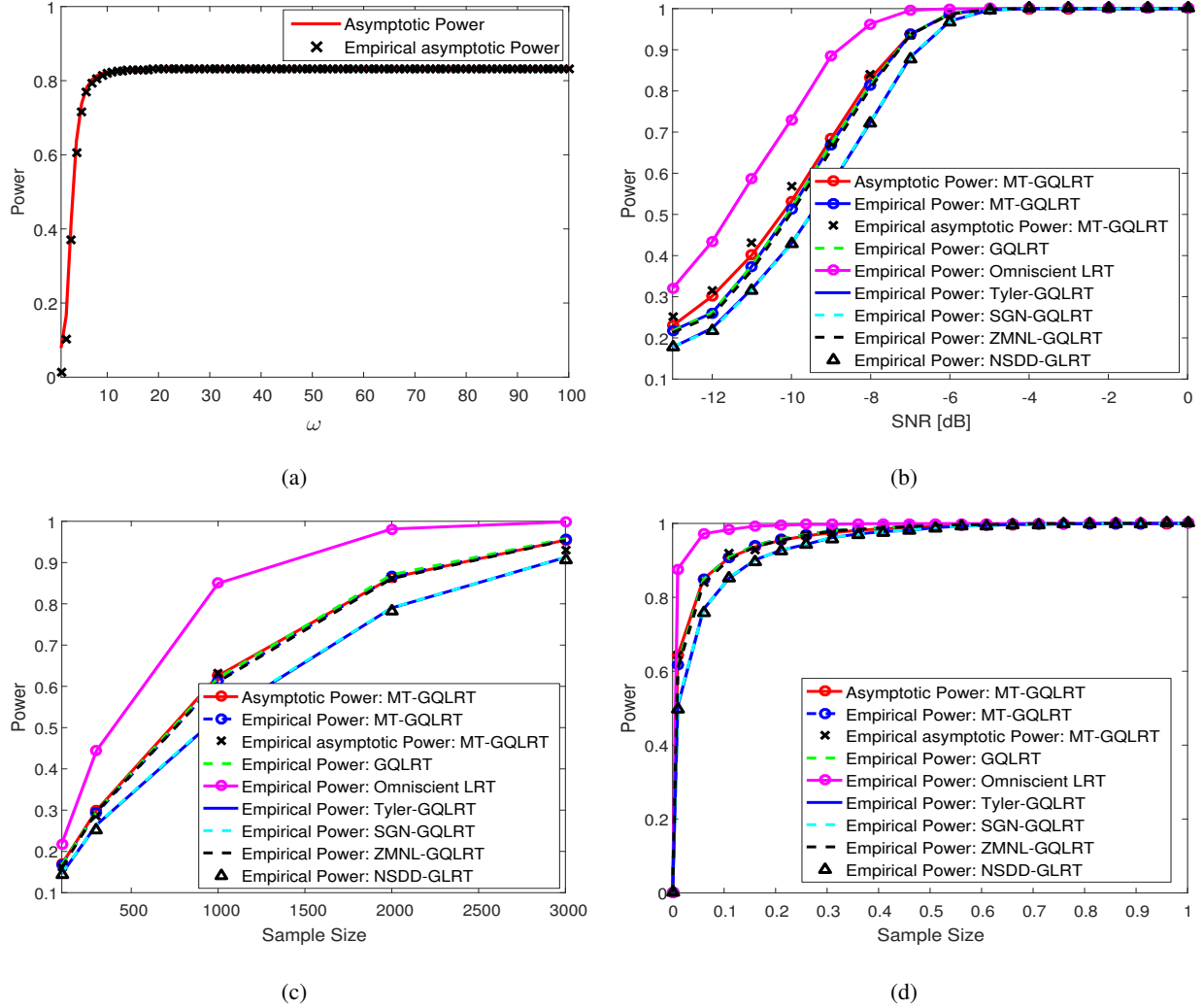


Fig. 1. **Signal detection in Gaussian noise:** (a) Asymptotic power (42) and its empirical estimate (43) versus the width parameter ω of the Gaussian MT-function (38). Notice that due to the consistency of (43), the compared quantities are close. (b) + (c) + (d) The empirical, asymptotic (42) and empirical asymptotic (43) power of the MT-GQLRT as a function of (b) SNR, (c) sample size and (d) test size as compared to the empirical power of the GQLRT, Tyler-GQLRT, SGN-GQLRT, ZMNL-GQLRT, NSDD-GLRT and the omniscient LRT. Notice that the MT-GQLRT outperforms the robust GQLRT alternatives when the noise is Gaussian.

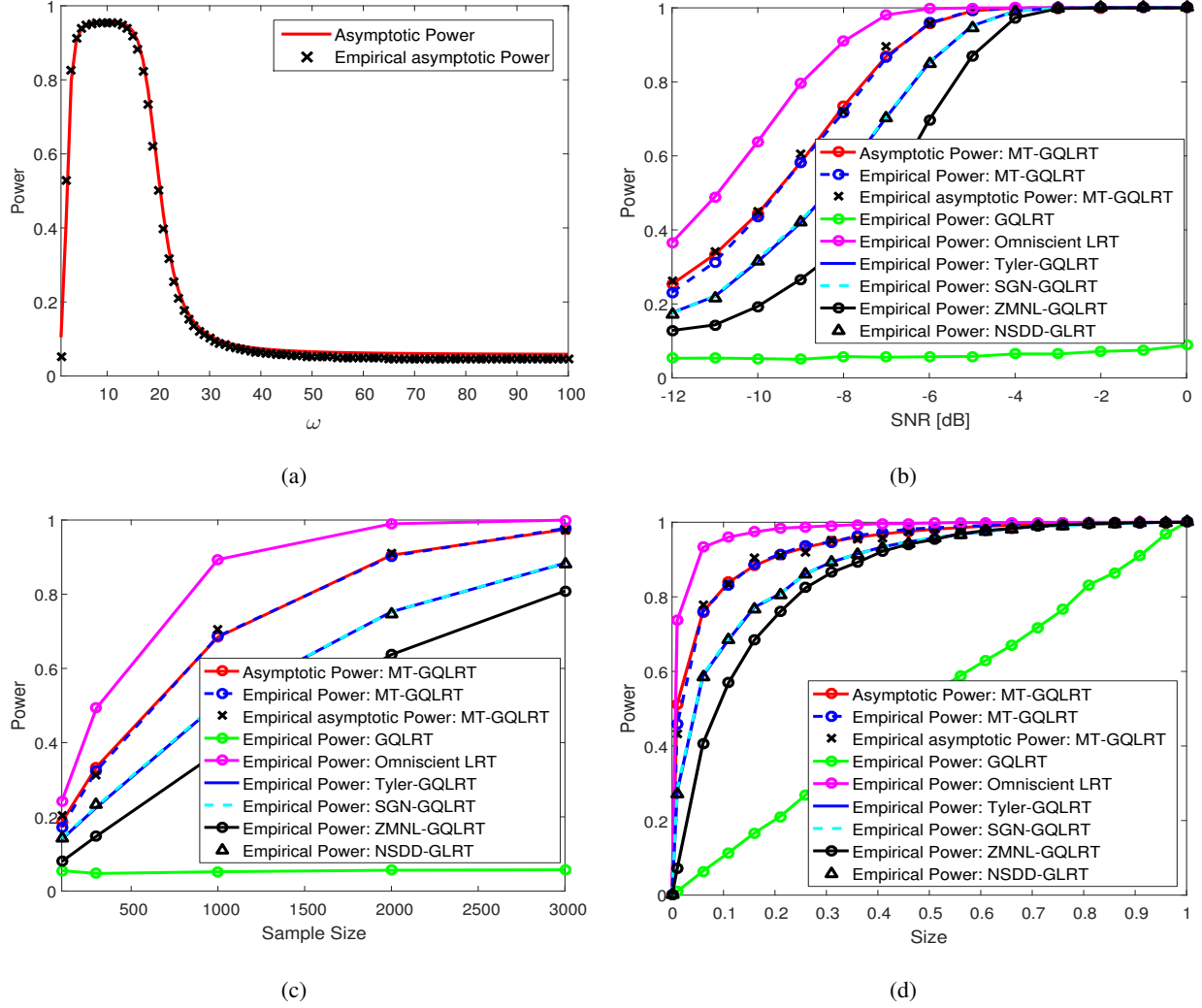


Fig. 2. **Signal detection in non-Gaussian noise:** (a) Asymptotic power (42) and its empirical estimate (43) versus the width parameter ω of the Gaussian MT-function (38). Notice that due to the consistency of (43), the compared quantities are close. (b) + (c) + (d) The empirical, asymptotic (42) and empirical asymptotic (43) power of the MT-GQLRT as a function of (b) SNR, (c) sample size and (d) size as compared to the empirical power of the GQLRT, Tyler-GQLRT, SGN-GQLRT, ZMNL-GQLRT, NSDD-GLRT and the omniscient LRT. One sees that the MT-GQLRT outperforms the GQLRT and its other robust alternatives and attains detection performance that are significantly closer to those obtained by the omniscient LRT.

B. Bayesian MT-GQLRT: Signal classification

We consider the following Bayesian signal classification problem:

$$\begin{aligned}
 H_0 &: \mathbf{X}_n = \mathbf{a}_0 + \mathbf{W}_n, \quad n = 1, \dots, N, \\
 H_1 &: \mathbf{X}_n = \mathbf{a}_1 + \mathbf{W}_n, \quad n = 1, \dots, N,
 \end{aligned} \tag{44}$$

with known a-priori probabilities P_{H_0} and P_{H_1} , where n is a discrete time index, $\mathbf{X}_n \in \mathbb{C}^p$ is an observation vector with $p > 2$ components, $\mathbf{a}_0, \mathbf{a}_1 \in \mathbb{C}^p$ are known deterministic vector signals, and $\mathbf{W}_n \in \mathbb{C}^p$ is a first-order stationary additive noise. Generally, this is a location parameter classification problem [50] when multiple instances from each class are available [50], [51].

In order to derive the MT-GQLRT for the Bayesian decision problem (44) we specify the MT-function in the set:

$$\left\{ u(\mathbf{x}) = v\left(\mathbf{P}_\mathbf{A}^\perp \mathbf{x}\right), \quad v: \mathbb{C}^p \rightarrow \mathbb{R}_+ \right\}, \quad (45)$$

where $\mathbf{A} \triangleq [\mathbf{a}_0, \mathbf{a}_1]$, $\mathbf{P}_\mathbf{A}^\perp \triangleq \mathbf{I}_p - \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H$ is the projection matrix into the subspace orthogonal to the span of \mathbf{a}_0 and \mathbf{a}_1 , and \mathbf{I}_p is a unit matrix of size of $p \times p$. Assuming that condition (5) is satisfied under the conditional probability measure $P_{\mathbf{x}|H}$, one can verify using (7), (9), (10), (44) and (45) that the conditional MT-mean and MT-covariance satisfy the following properties:

$$\boldsymbol{\mu}_{\mathbf{x}|H_k}^{(u)} = \mathbf{a}_k + \boldsymbol{\mu}_{\mathbf{W}}^{(u)} \quad k = 0, 1 \quad (46)$$

and

$$\boldsymbol{\Sigma}_{\mathbf{x}|H_k}^{(u)} = \boldsymbol{\Sigma}_{\mathbf{W}}^{(u)} \quad k = 0, 1 \quad (47)$$

where $\boldsymbol{\mu}_{\mathbf{W}}^{(u)}$ and $\boldsymbol{\Sigma}_{\mathbf{W}}^{(u)}$ are the MT-mean and MT-covariance of the noise component. Hence, by substituting (46) and (47) into (27) the resulting test statistic after subtraction of the observation-independent constant $c_1^{(u)} \triangleq \boldsymbol{\mu}_{\mathbf{x}|H_0}^{(u)H} \left(\boldsymbol{\Sigma}_{\mathbf{W}}^{(u)}\right)^{-1} \boldsymbol{\mu}_{\mathbf{x}|H_0}^{(u)} - \boldsymbol{\mu}_{\mathbf{x}|H_1}^{(u)H} \left(\boldsymbol{\Sigma}_{\mathbf{W}}^{(u)}\right)^{-1} \boldsymbol{\mu}_{\mathbf{x}|H_1}^{(u)}$ is given by:

$$S'_u \triangleq S_u - c_1^{(u)} = 2 \operatorname{Re} \left\{ (\mathbf{a}_1 - \mathbf{a}_0)^H \left(\boldsymbol{\Sigma}_{\mathbf{W}}^{(u)}\right)^{-1} \hat{\boldsymbol{\mu}}_{\mathbf{x}}^{(u)} \right\}. \quad (48)$$

Similarly to the detection problem in the previous subsection, we further assume that the noise component has a density that is spherically contoured with the stochastic representation (37). Again, in order to mitigate the effect of outliers and involve higher-order statistical moments, we specify the MT-function in a subset of (45) that is comprised of zero-centred Gaussian functions parametrized by a width parameter ω , i.e.,

$$u_G(\mathbf{x}; \omega) = \exp \left(-\left\| \mathbf{P}_\mathbf{A}^\perp \mathbf{x} \right\|^2 / \omega^2 \right), \quad \omega \in \mathbb{R}_{++}. \quad (49)$$

Similarly to the signal detection example, it can be shown that the resulting empirical MT-mean and MT-covariance that comprise the test-statistic are robust to outliers with sufficiently high probability. Using (9), (10), (37) and (49) it can be shown that the MT-covariance of the noise satisfy:

$$\boldsymbol{\Sigma}_{\mathbf{W}}^{(u_G)}(\omega) = r_0(\omega) \mathbf{P}_\mathbf{A} + r_1(\omega) \mathbf{I}, \quad (50)$$

where $\mathbf{P}_\mathbf{A}$ is the projection matrix onto the column space of \mathbf{A} , and $r_0(\omega)$ and $r_1(\omega)$ are some strictly positive functions of ω . Hence, by substituting (50) into (48) followed by normalization by the observation-independent factor $c_2(\omega) \triangleq 2/(r_0(\omega) + r_1(\omega))$, the Bayesian MT-GQLRT (19) simplifies to:

$$S''_{u_G} \triangleq S'_{u_G}/c_2(\omega) = \text{Re} \left\{ (\mathbf{a}_1 - \mathbf{a}_0)^H \hat{\boldsymbol{\mu}}_{\mathbf{x}}^{(u_G)}(\omega) \right\} \underset{H_0}{\overset{H_1}{\gtrless}} t'', \quad (51)$$

where $t'' \triangleq (t - c_1^{(u_G)})/c_2(\omega)$.

We choose the loss coefficients $L_{10} = L_{01} = 1$, under which the asymptotic Bayes risk (29) reduces to the probability of error [1]. In this case it can be shown that the asymptotic minimum probability of error w.r.t. the threshold parameter takes the form:

$$P_e^{(u_G)}(\omega) = \sum_{k=0}^1 P_{H_k} Q \left(G(\omega) + (-1)^k \frac{1}{2G(\omega)} \log \frac{P_{H_0}}{P_{H_1}} \right), \quad (52)$$

where $G(\omega) \triangleq \frac{\|\mathbf{a}_1 - \mathbf{a}_0\| \text{E} \left[\left(\frac{\omega^2}{\omega^2 + \tilde{\nu}^2} \right)^{p-2}; P_\nu \right]}{\sqrt{\frac{2}{N} \text{E} \left[\tilde{\nu}^2 \left(\frac{\omega^2}{\omega^2 + 2\tilde{\nu}^2} \right)^{p-2}; P_\nu \right]}}$ and $\tilde{\nu} \triangleq \sigma_{\mathbf{z}}^2 \nu$. Moreover, by (30) and (31) the empirical estimate of (52) is given by:

$$\hat{P}_e^{(u_G)}(\omega) = \sum_{k=0}^1 P_{H_k} Q \left(\frac{\tilde{t}_{opt}^{(u_G)}(\omega) - \tilde{\kappa}_{H_k}^{(u_G)}(\omega)}{\sqrt{\hat{\gamma}_{H_k}^{(u_G)}(\omega)}} \right), \quad (53)$$

where

$$\begin{aligned} \tilde{\kappa}_{H_k}^{(u_G)}(\omega) &\triangleq \frac{\hat{\kappa}_{H_k}^{(u_G)}(\omega) - c_1^{(u_G)}}{c_2(\omega)} = \sum_{n=1}^N \hat{\varphi}_{u_G}(\mathbf{X}_n^{(k)}; \omega) \text{Re} \left\{ (\mathbf{a}_1 - \mathbf{a}_0)^H \mathbf{X}_n^{(k)} \right\}, \\ \hat{\gamma}_{H_k}^{(u_G)}(\omega) &\triangleq \frac{\hat{\gamma}_{H_k}^{(u_G)}(\omega)}{c_2^2(\omega)} = \frac{N_k}{N} \sum_{n=1}^{N_k} \hat{\varphi}_{u_G}^2(\mathbf{X}_n^{(k)}; \omega) \left(\text{Re} \left\{ (\mathbf{a}_1 - \mathbf{a}_0)^H \mathbf{X}_n^{(k)} \right\} - \tilde{\kappa}_{H_k}^{(u_G)}(\omega) \right)^2, \end{aligned}$$

$k = 0, 1$, and the optimal threshold $\tilde{t}_{opt}^{(u_G)}(\omega)$ is obtained from (31) by replacing $\hat{\kappa}_{H_k}^{(u_G)}$ and $\hat{\gamma}_{H_k}^{(u_G)}$ with $\tilde{\kappa}_{H_k}^{(u_G)}(\omega)$ and $\tilde{\gamma}_{H_k}^{(u_G)}(\omega)$. As discussed in Subsection V-C, the empirical error probability (53) will be used for optimal choice of the width parameter ω of the Gaussian MT-function (49).

In the following simulation examples we compare the classification performance of the MT-GQLRT (51) to the Bayesian versions of the omniscient LRT, the standard GQLRT and other robust GQLRT extensions. Under the classification problem (44) one can verify that the test-statistic of the GQLRT reduces to $S_{GQLRT} = \text{Re} \left\{ (\mathbf{a}_1 - \mathbf{a}_0)^H \hat{\boldsymbol{\mu}}_{\mathbf{x}} \right\}$, where $\hat{\boldsymbol{\mu}}_{\mathbf{x}}$ is the standard SMV. Hence, other robust alternatives to the GQLRT can be obtained by replacing the non-robust SMV with robust location estimates, namely, the median estimator, and Tukey's bi-square M-estimator [52] whose implementation is described in appendix F. These robust GQLRT extensions are called here, Median-GQLRT and Tukey-GQLRT.

In all examples, the vectors \mathbf{a}_0 and \mathbf{a}_1 were set to $\mathbf{a}_k \triangleq s_k [1, e^{-i\pi/p}, \dots, e^{-i\pi(p-1)/p}]^T$, $k = 0, 1$, where $s_0 = 5$, $s_1 = 5.25$ and $p = 10$. The a-priori probabilities were set to $P_{H_0} = 0.6$ and $P_{H_1} = 0.4$. We considered two types of noise distributions with zero location parameter and isotropic dispersion $\sigma_Z^2 \mathbf{I}_p$: 1) Gaussian and 2) t -distributed noise [41] with $\lambda = 0.2$ degrees of freedom.

Similarly to the signal detection example, for each noise type we performed two simulations. In the first simulation example, we compared the asymptotic probability of error (52) to its empirical estimate (53) as a function of ω for sample size of $N = 300$. The empirical asymptotic probability of error (53) was obtained using two i.i.d. training sequences from $P_{\mathbf{x}|H_0}$ and $P_{\mathbf{x}|H_1}$ containing $N_0 = N_1 = 3 \times 10^4$ samples. The SNR, defined in this example as $\text{SNR} \triangleq 10 \log_{10} (\|\mathbf{a}_0 - \mathbf{a}_1\|)^2 / \sigma_Z^2$ was set to -18 [dB]. Observing Figs. 3(a) and 4(a), one sees that due to the consistency of (53) the compared quantities are very close. This illustrates the reliability of the empirical asymptotic Bayes risk in optimal choice of the of the MT-function parameter, as discussed in subsection V-C.

In the second example, illustrated in Figs. 3(b), 3(c), 4(b) and 4(c), we compared the empirical, asymptotic (52) and empirical asymptotic (53) probability of error of the MT-GQLRT to the empirical probability of error of the Bayesian GQLRT, Tukey-GQLRT, Median-GQLRT, and the omniscient Bayesian LRT. Note that in the Gaussian noise case the classification performance of the GQLRT are omitted from the figures as under (44) it must coincide with the LRT. The optimal Gaussian MT-function parameter ω_{opt} was obtained by minimizing (53) over $\Omega = [1, 100]$. The empirical probability of error curves were obtained using 10^5 Monte-Carlo simulations. The SNR and sample size are used to index the classification performances as depicted in Figs. 3(b) and 3(c) for the Gaussian noise and in Figs. 4(b) and 4(c) for the t -distributed noise. The probability of error versus SNR was evaluated for $N = 300$ i.i.d observations, and the probability of error versus sample size was evaluated for $\text{SNR} = -22$ [dB]. Observing Figs. 3(b) and 3(c), one can notice that, as expected, when the noise is Gaussian, the MT-GQLRT achieves the LRT performance and outperforms the Median-GQLRT and Tukey-GQLRT. Observing Figs. 4(b) and 4(c), one sees that for the t -distributed noise, the MT-GQLRT outperforms the non-robust Bayesian GQLRT and its other robust generalizations and attains classification performance that are much closer to those obtained by the Bayesian LRT that, unlike the MT-GQLRT, requires complete knowledge of the conditional likelihood function under each hypothesis.

VII. CONCLUSION

In this paper a new test, called MT-GQLRT, for non-Bayesian binary hypothesis testing was developed that is based on the GQLRT after transformation of the probability distribution of the data. A Bayesian

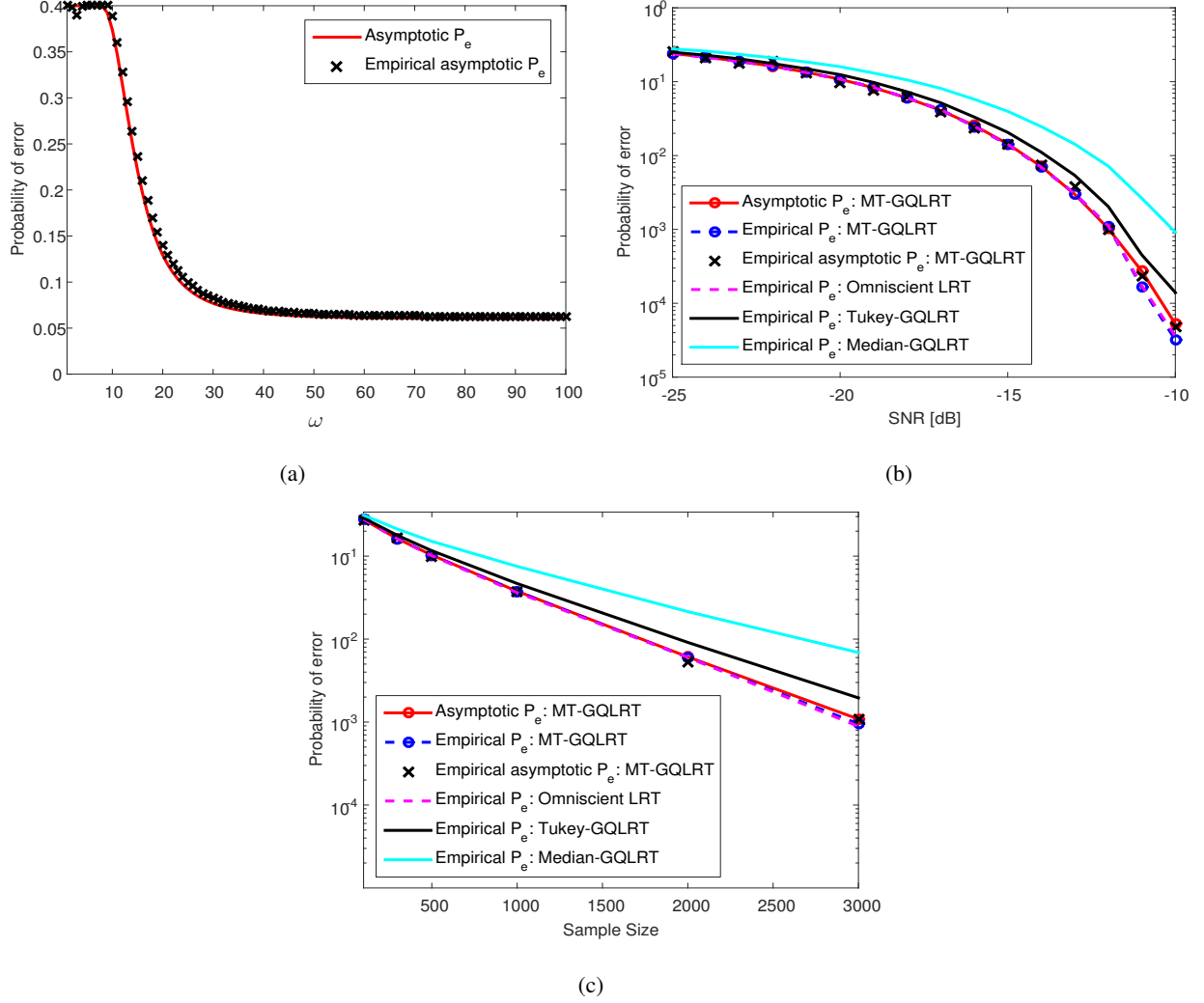


Fig. 3. **Signal classification in Gaussian noise:** (a) Asymptotic probability of error (52) and its empirical estimate (53) versus the width parameter ω of the Gaussian MT-function (49). Notice that due to consistency of (53) the compared quantities are close. (b) + (c) The empirical, asymptotic (52) and empirical asymptotic (53) probability of error of the MT-GQLRT as a function of (b) SNR and (c) sample size as compared to the empirical probability of error of the Tukey-GQLRT, Median-GQLRT and the omniscient LRT. Notice that the MT-GQLRT outperforms the robust GQLRT alternatives and attains the classification performance of the LRT.

extension of this test is also developed that applies the transformation to the conditional probability distribution of the data. By specifying the MT-function in the Gaussian family of functions the non-Bayesian and Bayesian MT-GQLRTs were applied to robust signal detection and classification, respectively. This paper has demonstrated yet another instance where the powerful measure transformation approach can result in a significant performance boost of classical statistical signal processing algorithms, adding robust hypothesis testing to a growing list of applications that has previously included canonical correlation

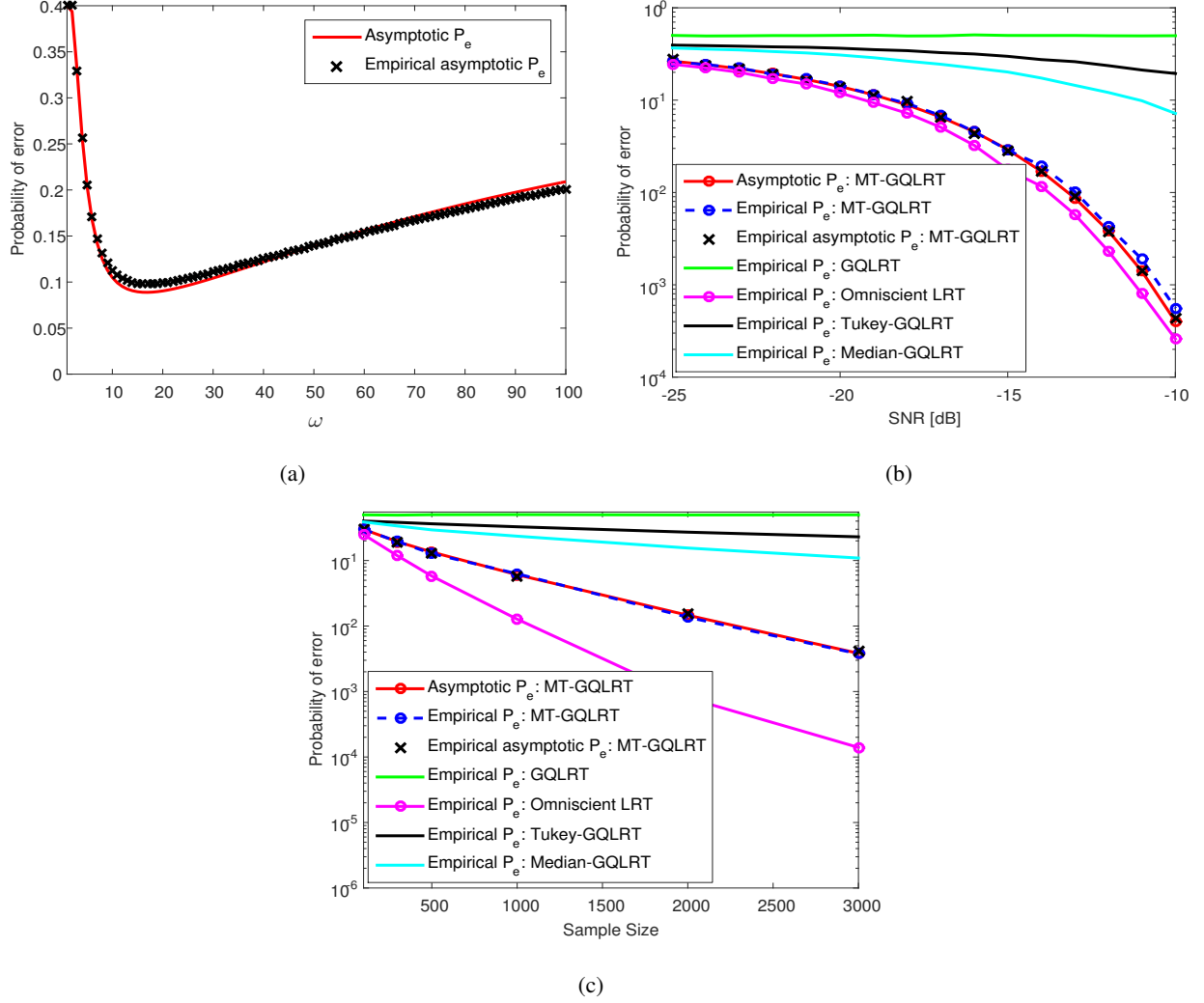


Fig. 4. **Signal classification in non-Gaussian noise:** (a) Asymptotic probability of error (52) and its empirical estimate (53) versus the width parameter ω of the Gaussian MT-function (49). Notice that due to consistency of (53) the compared quantities are close. (b) + (c) The empirical, asymptotic (52) and empirical asymptotic (53) probability of error of the MT-GQLRT as a function of (b) SNR and (c) sample size as compared to the empirical probability of error of the GQLRT, Tukey-GQLRT, Median-GQLRT and the omniscient LRT. Notice that the MT-GQLRT outperforms the GQLRT and its other robust generalizations and significantly reduces the performance gap relative to the LRT.

analysis [15], [16], multiple signal classification (MUSIC) [17], [18] and parameter estimation [19], [20]. We expect that many other applications can also benefit from measure transformation approaches.

APPENDIX

A. Proof of Theorem 1:

: By (13), (16), (17), (18) and assumption A-1, the test statistic is a non-degenerate random variable that can be written as:

$$T_u = \frac{\frac{1}{N} \sum_{n=1}^N u(\mathbf{X}_n) \psi_u(\mathbf{X}_n)}{\frac{1}{N} \sum_{n=1}^N u(\mathbf{X}_n)}. \quad (54)$$

Since $\mathbf{X}_n, n = 1, \dots, N$ are i.i.d random vectors and the functions $u(\cdot)$ and $\psi_u(\cdot)$ are real, the products $u(\mathbf{X}_n) \psi_u(\mathbf{X}_n), n = 1, \dots, N$ are i.i.d and real. According to (5), (7) and (20) $u(\mathbf{X}) \left(\psi_u(\mathbf{X}) - \eta_H^{(u)} \right)$ is a zero-mean random variable under $P_{\mathbf{X};H}$ for any $H \in \mathcal{H}$. Furthermore, by assumptions A-2, A-3 and Lemma 1 stated in Appendix B, its variance under $P_{\mathbf{X};H}$ is finite for any $H \in \mathcal{H}$. Therefore, by the central limit theorem [53] we conclude that the translated and scaled version of the nominator in (54) satisfies:

$$\sqrt{\frac{N}{\tilde{\lambda}_H^{(u)}}} \frac{1}{N} \sum_{n=1}^N u(\mathbf{X}_n) \left(\psi_u(\mathbf{X}_n) - \eta_H^{(u)} \right) \xrightarrow[N \rightarrow \infty]{D} \mathcal{N}(0, 1) \quad \forall H \in \mathcal{H}, \quad (55)$$

where

$$\tilde{\lambda}_H^{(u)} \triangleq \mathbb{E} \left[u^2(\mathbf{X}) \left(\psi_u(\mathbf{X}) - \eta_H^{(u)} \right)^2; P_{\mathbf{X};H} \right]. \quad (56)$$

Since by Definition 1 $u(\mathbf{X})$ is non-negative and $0 < \mathbb{E}[u(\mathbf{X}); P_{\mathbf{X};H}] < \infty$ for $H = H_0$ and $H = H_1$, by Khinchine's strong law of large numbers [33] we have that the denominator in (54) satisfies:

$$\frac{1}{N} \sum_{n=1}^N u(\mathbf{X}_n) \xrightarrow[N \rightarrow \infty]{w.p.1} \mathbb{E}[u(\mathbf{X}); P_{\mathbf{X};H}] \quad \forall H \in \mathcal{H}. \quad (57)$$

Notice that by (7), (21) and (56), $\lambda_H^{(u)} = \frac{\tilde{\lambda}_H^{(u)}}{N \mathbb{E}^2[u(\mathbf{X}); P_{\mathbf{X};H}]}$. Therefore, by (54)-(57) and Slutsky's theorem [36]:

$$\frac{T_u - \eta_H^{(u)}}{\sqrt{\lambda_H^{(u)}}} = \frac{\sqrt{\frac{N}{\tilde{\lambda}_H^{(u)}}} \frac{1}{N} \sum_{n=1}^N u(\mathbf{X}_n) \left(\psi_u(\mathbf{X}_n) - \eta_H^{(u)} \right)}{\left(\frac{1}{N} \sum_{n=1}^N u(\mathbf{X}_n) \right) / \mathbb{E}[u(\mathbf{X}); P_{\mathbf{X};H}]} \xrightarrow[N \rightarrow \infty]{D} \mathcal{N}(0, 1) \quad \forall H \in \mathcal{H}.$$

■

B. Lemma:

Lemma 1. Assume that $\Sigma_{\mathbf{X};H}^{(u)}$ is non-singular, $\mathbb{E}[u^2(\mathbf{X}); P_{\mathbf{X};H}]$ and $\mathbb{E}[\|\mathbf{X}\|^4 u^2(\mathbf{X}); P_{\mathbf{X};H}]$ are finite for $H = H_0$ and $H = H_1$. Then, the expectations $A \triangleq \mathbb{E}[u(\mathbf{X}) |\psi_u(\mathbf{X})|; P_{\mathbf{X};H}]$ and $B \triangleq \mathbb{E} \left[u^2(\mathbf{X}) \left(\psi_u(\mathbf{X}) - \eta_H^{(u)} \right)^2; P_{\mathbf{X};H} \right]$ are finite for any $H \in \mathcal{H}$.

Proof: By (16), (18), the non-singularity of $\Sigma_{\mathbf{x};H_0}^{(u)}$ and $\Sigma_{\mathbf{x};H_1}^{(u)}$, inequality (1) in [54], and the triangle inequality:

$$\begin{aligned}
|\psi_u(\mathbf{X})| &\leq \left| \log \frac{\det \Sigma_{\mathbf{x};H_0}^{(u)}}{\det \Sigma_{\mathbf{x};H_1}^{(u)}} \right| + \left\| \Sigma_{\mathbf{x};H_0}^{(u)-1/2} (\mathbf{X} - \boldsymbol{\mu}_{\mathbf{x};H_0}^{(u)}) \right\|^2 + \left\| \Sigma_{\mathbf{x};H_1}^{(u)-1/2} (\mathbf{X} - \boldsymbol{\mu}_{\mathbf{x};H_1}^{(u)}) \right\|^2 \\
&\leq \left| \log \frac{\det \Sigma_{\mathbf{x};H_0}^{(u)}}{\det \Sigma_{\mathbf{x};H_1}^{(u)}} \right| + \left\| \Sigma_{\mathbf{x};H_0}^{(u)-1/2} \right\|_S^2 \left\| \mathbf{X} - \boldsymbol{\mu}_{\mathbf{x};H_0}^{(u)} \right\|^2 + \left\| \Sigma_{\mathbf{x};H_1}^{(u)-1/2} \right\|_S^2 \left\| \mathbf{X} - \boldsymbol{\mu}_{\mathbf{x};H_1}^{(u)} \right\|^2 \\
&= \left| \log \frac{\det \Sigma_{\mathbf{x};H_0}^{(u)}}{\det \Sigma_{\mathbf{x};H_1}^{(u)}} \right| + \lambda_{\min}^{-1}(\Sigma_{\mathbf{x};H_0}^{(u)}) \left\| \mathbf{X} - \boldsymbol{\mu}_{\mathbf{x};H_0}^{(u)} \right\|^2 + \lambda_{\min}^{-1}(\Sigma_{\mathbf{x};H_1}^{(u)}) \left\| \mathbf{X} - \boldsymbol{\mu}_{\mathbf{x};H_1}^{(u)} \right\|^2 \\
&\leq \left| \log \frac{\det \Sigma_{\mathbf{x};H_0}^{(u)}}{\det \Sigma_{\mathbf{x};H_1}^{(u)}} \right| + \lambda_{\min}^{-1}(\Sigma_{\mathbf{x};H_0}^{(u)}) \left(\left\| \mathbf{X} \right\| + \left\| \boldsymbol{\mu}_{\mathbf{x};H_0}^{(u)} \right\| \right)^2 + \lambda_{\min}^{-1}(\Sigma_{\mathbf{x};H_1}^{(u)}) \left(\left\| \mathbf{X} \right\| + \left\| \boldsymbol{\mu}_{\mathbf{x};H_1}^{(u)} \right\| \right)^2 \\
&= c_1 \left\| \mathbf{X} \right\|^2 + 2c_2 \left\| \mathbf{X} \right\| + c_3,
\end{aligned} \tag{58}$$

where $\|\cdot\|_S$ denote the spectral norm, $\lambda_{\min}(\cdot)$ denote the minimal eigenvalue of a matrix and $c_1 \triangleq \lambda_{\min}^{-1}(\Sigma_{\mathbf{x};H_0}^{(u)}) + \lambda_{\min}^{-1}(\Sigma_{\mathbf{x};H_1}^{(u)})$, $c_2 \triangleq \left\| \boldsymbol{\mu}_{\mathbf{x};H_0}^{(u)} \right\| \lambda_{\min}^{-1}(\Sigma_{\mathbf{x};H_0}^{(u)}) + \left\| \boldsymbol{\mu}_{\mathbf{x};H_1}^{(u)} \right\| \lambda_{\min}^{-1}(\Sigma_{\mathbf{x};H_1}^{(u)})$, $c_3 \triangleq \left\| \boldsymbol{\mu}_{\mathbf{x};H_0}^{(u)} \right\|^2 \lambda_{\min}^{-1}(\Sigma_{\mathbf{x};H_0}^{(u)}) + \left\| \boldsymbol{\mu}_{\mathbf{x};H_1}^{(u)} \right\|^2 \lambda_{\min}^{-1}(\Sigma_{\mathbf{x};H_1}^{(u)}) + \left| \log \frac{\det \Sigma_{\mathbf{x};H_0}^{(u)}}{\det \Sigma_{\mathbf{x};H_1}^{(u)}} \right|$.

By Definition 1, Hölder's inequality [36] and the assumption that $E[u^2(\mathbf{X}); P_{\mathbf{x};H}]$ and $E[\left\| \mathbf{X} \right\|^4 u^2(\mathbf{X}); P_{\mathbf{x};H}]$ are finite for $H = H_0$ and $H = H_1$:

$$E[u(\mathbf{X}); P_{\mathbf{x};H}] < \infty \tag{59a}$$

$$E[u^2(\mathbf{X}) \left\| \mathbf{X} \right\|^2; P_{\mathbf{x};H}] \leq \sqrt{E[u^2(\mathbf{X}) \left\| \mathbf{X} \right\|^4; P_{\mathbf{x};H}] E[u^2(\mathbf{X}); P_{\mathbf{x};H}]} < \infty \tag{59b}$$

$$E[u^2(\mathbf{X}) \left\| \mathbf{X} \right\|^3; P_{\mathbf{x};H}] \leq \sqrt{E[u^2(\mathbf{X}) \left\| \mathbf{X} \right\|^2; P_{\mathbf{x};H}] E[u^2(\mathbf{X}) \left\| \mathbf{X} \right\|^4; P_{\mathbf{x};H}]} < \infty \tag{59c}$$

$$E[u^2(\mathbf{X}) \left\| \mathbf{X} \right\|; P_{\mathbf{x};H}] \leq \sqrt{E[u^2(\mathbf{X}); P_{\mathbf{x};H}] E[u^2(\mathbf{X}) \left\| \mathbf{X} \right\|^2; P_{\mathbf{x};H}]} < \infty \tag{59d}$$

for all $H \in \mathcal{H}$. Therefore, by (58), (59) and Hölder's inequality [36]:

$$\begin{aligned}
A &\triangleq E[u(\mathbf{X}) |\psi_u(\mathbf{X})|; P_{\mathbf{x};H}] \\
&\leq E\left[u(\mathbf{X}) (c_1 \left\| \mathbf{X} \right\|^2 + 2c_2 \left\| \mathbf{X} \right\| + c_3); P_{\mathbf{x};H}\right] \\
&\leq \sqrt{E\left[u^2(\mathbf{X}) (c_1 \left\| \mathbf{X} \right\|^2 + 2c_2 \left\| \mathbf{X} \right\| + c_3)^2; P_{\mathbf{x};H}\right]} < \infty \quad \forall H \in \mathcal{H}.
\end{aligned} \tag{60}$$

According to (5), (7), (20) and (60) it follows that $\left| \eta_H^{(u)} \right|$ is finite since $\left| \eta_H^{(u)} \right| \leq \frac{A}{E[u(\mathbf{X}); P_{\mathbf{x};H}]} < \infty$.

Moreover, by (58)

$$\begin{aligned}
(\psi_u(\mathbf{X}) - \eta_H^{(u)})^2 &\leq \psi_u^2(\mathbf{X}) + 2 \left| \psi_u(\mathbf{X}) \eta_H^{(u)} \right| + (\eta_H^{(u)})^2 \\
&\leq (c_1 \left\| \mathbf{X} \right\|^2 + 2c_2 \left\| \mathbf{X} \right\| + c_3)^2 + 2 \left| \eta_H^{(u)} \right| (c_1 \left\| \mathbf{X} \right\|^2 + 2c_2 \left\| \mathbf{X} \right\| + c_3) + (\eta_H^{(u)})^2 \\
&= \tilde{c}_4 \left\| \mathbf{X} \right\|^4 + \tilde{c}_3 \left\| \mathbf{X} \right\|^3 + \tilde{c}_2 \left\| \mathbf{X} \right\|^2 + \tilde{c}_1 \left\| \mathbf{X} \right\| + \tilde{c}_0
\end{aligned} \tag{61}$$

for all $H \in \mathcal{H}$, where $\tilde{c}_4 \triangleq c_1^2$, $\tilde{c}_3 \triangleq 2c_1c_2$, $\tilde{c}_2 \triangleq 4c_2^2 + 2\left|\eta_H^{(u)}\right|c_1 + 2c_1c_3$, $\tilde{c}_1 \triangleq 2c_2c_3 + 4\left|\eta_H^{(u)}\right|c_2$, and $\tilde{c}_0 \triangleq c_3^2 + \left(\eta_H^{(u)}\right)^2 + 2\left|\eta_H^{(u)}\right|c_3$. Finally, by (59), (61) and the fact that $\left|\eta_H^{(u)}\right|$ is finite we conclude that:

$$B \leq \mathbb{E} \left[u^2(\mathbf{X}) \left(\tilde{c}_4 \|\mathbf{X}\|^4 + \tilde{c}_3 \|\mathbf{X}\|^3 + \tilde{c}_2 \|\mathbf{X}\|^2 + \tilde{c}_1 \|\mathbf{X}\| + \tilde{c}_0 \right); P_{\mathbf{X};H} \right] < \infty \quad \forall H \in \mathcal{H}.$$

■

C. Proof of proposition 2:

: By (13), (18), (24), (25) and assumptions A-1, A-2, the empirical estimators $\hat{\eta}_{H_k}^{(u)}$ and $\hat{\lambda}_{H_k}^{(u)}$ are non-degenerate random variables that can be written as:

$$\hat{\eta}_{H_k}^{(u)} \triangleq \frac{\frac{1}{N_k} \sum_{n=1}^{N_k} u(\mathbf{X}_n^{(k)}) \psi_u(\mathbf{X}_n^{(k)})}{\frac{1}{N_k} \sum_{n=1}^{N_k} u(\mathbf{X}_n^{(k)})} \quad (62)$$

and

$$\hat{\lambda}_{H_k}^{(u)} \triangleq \frac{1}{N} \frac{\frac{1}{N_k} \sum_{n=1}^{N_k} u^2(\mathbf{X}_n^{(k)}) \left(\psi_u^2(\mathbf{X}_n^{(k)}) - 2\psi_u(\mathbf{X}_n^{(k)}) \hat{\eta}_{H_k}^{(u)} + \left(\hat{\eta}_{H_k}^{(u)} \right)^2 \right)}{\left(\frac{1}{N_k} \sum_{n=1}^{N_k} u(\mathbf{X}_n^{(k)}) \right)^2}, \quad (63)$$

respectively. Since $\{\mathbf{X}_n^{(k)}\}_{n=1}^{N_k}$ are i.i.d and the functions $u(\cdot)$ and $\psi_u(\cdot)$ are real, the products $\left\{ u(\mathbf{X}_n^{(k)}) \psi_u(\mathbf{X}_n^{(k)}) \right\}_{n=1}^{N_k}$, $\left\{ u^2(\mathbf{X}_n^{(k)}) \psi_u^2(\mathbf{X}_n^{(k)}) \right\}_{n=1}^{N_k}$ and $\left\{ u^2(\mathbf{X}_n^{(k)}) \psi_u(\mathbf{X}_n^{(k)}) \right\}_{n=1}^{N_k}$ define real i.i.d. sequences.

Furthermore, by Hölder's inequality [36], assumptions A-2, A-3 and Lemma 1 stated in Appendix B we have that $\mathbb{E}[u(\mathbf{X}); P_{\mathbf{X};H_k}] < \infty$, $\mathbb{E}[u^2(\mathbf{X}); P_{\mathbf{X};H_k}] < \infty$, $\mathbb{E}[u(\mathbf{X})|\psi_u(\mathbf{X})|; P_{\mathbf{X};H_k}] < \infty$, $\mathbb{E}[u^2(\mathbf{X})\psi_u^2(\mathbf{X}); P_{\mathbf{X};H_k}] < \infty$ and $\mathbb{E}[u^2(\mathbf{X})|\psi_u(\mathbf{X})|; P_{\mathbf{X};H_k}] < \infty$ for any $k \in \{0, 1\}$. Therefore, by Khinchine's strong law of large numbers [33]:

$$\frac{1}{N_k} \sum_{n=1}^{N_k} u(\mathbf{X}_n^{(k)}) \xrightarrow[N_k \rightarrow \infty]{w.p. 1} \mathbb{E}[u(\mathbf{X}); P_{\mathbf{X};H_k}], \quad (64)$$

$$\frac{1}{N_k} \sum_{n=1}^{N_k} u^2(\mathbf{X}_n^{(k)}) \xrightarrow[N_k \rightarrow \infty]{w.p. 1} \mathbb{E}[u^2(\mathbf{X}); P_{\mathbf{X};H_k}], \quad (65)$$

$$\frac{1}{N_k} \sum_{n=1}^{N_k} u(\mathbf{X}_n^{(k)}) \psi_u(\mathbf{X}_n^{(k)}) \xrightarrow[N_k \rightarrow \infty]{w.p. 1} \mathbb{E}[u(\mathbf{X})\psi_u(\mathbf{X}); P_{\mathbf{X};H_k}], \quad (66)$$

$$\frac{1}{N_k} \sum_{n=1}^{N_k} u^2(\mathbf{X}_n^{(k)}) \psi_u^2(\mathbf{X}_n^{(k)}) \xrightarrow[N_k \rightarrow \infty]{w.p. 1} \mathbb{E}[u^2(\mathbf{X})\psi_u^2(\mathbf{X}); P_{\mathbf{X};H_k}] \quad (67)$$

and

$$\frac{1}{N_k} \sum_{n=1}^{N_k} u^2(\mathbf{X}_n^{(k)}) \psi_u(\mathbf{X}_n^{(k)}) \xrightarrow[N_k \rightarrow \infty]{w.p. 1} \mathbb{E}[u^2(\mathbf{X})\psi_u(\mathbf{X}); P_{\mathbf{X};H_k}] \quad (68)$$

for any $k \in \{0, 1\}$. Hence, by (7), (20), (21), (62)-(68) and Mann-Wald's Theorem [55] we conclude that

$$\hat{\eta}_{H_k}^{(u)} \xrightarrow[N_k \rightarrow \infty]{w.p. \, 1} \eta_{H_k}^{(u)} \quad \text{and} \quad \hat{\lambda}_{H_k}^{(u)} \xrightarrow[N_k \rightarrow \infty]{w.p. \, 1} \lambda_{H_k}^{(u)}, \quad k = 0, 1. \quad (69)$$

Therefore, by (22), (23), (69), the continuity of the standard normal tail probability $Q(\cdot)$ and Mann-Wald's Theorem [55] we conclude that $\hat{\alpha}_u \xrightarrow[N_0 \rightarrow \infty]{w.p. \, 1} \alpha_u$ and $\hat{\beta}_u \xrightarrow[N_1 \rightarrow \infty]{w.p. \, 1} \beta_u$. ■

D. Proof of proposition 4:

: Similarly to the proof of Proposition 2 stated in Appendix C, one can verify that under conditions B-1 - B-3

$$\hat{\kappa}_{H_k}^{(u)} \xrightarrow[N_k \rightarrow \infty]{w.p. \, 1} \kappa_{H_k}^{(u)} \quad \text{and} \quad \hat{\gamma}_{H_k}^{(u)} \xrightarrow[N_k \rightarrow \infty]{w.p. \, 1} \gamma_{H_k}^{(u)} \quad (70)$$

for any $k \in \{0, 1\}$. Therefore, by (29), (30), (70), the continuity of $Q(\cdot)$ and Mann-Wald's Theorem [55] we conclude that $\hat{R}^{(u)} \xrightarrow[N_0, N_1 \rightarrow \infty]{w.p. \, 1} R^{(u)}$. ■

E. Proof of proposition 5:

: One can verify that if assumption C-1 is satisfied and $\hat{\gamma}_{H_1}^{(u)} \neq \hat{\gamma}_{H_1}^{(u)}$, then the only two stationary points [56] of $\hat{R}^{(u)}(\cdot)$ are given by

$$t_1^* \triangleq \frac{\hat{\gamma}_{H_0}^{(u)} \hat{\kappa}_{H_1}^{(u)} - \hat{\gamma}_{H_1}^{(u)} \hat{\kappa}_{H_0}^{(u)} - \sqrt{\hat{\gamma}_{H_0}^{(u)} \hat{\gamma}_{H_1}^{(u)} \hat{s}^{(u)}}}{\hat{\gamma}_{H_0}^{(u)} - \hat{\gamma}_{H_1}^{(u)}}$$

and

$$t_2^* \triangleq \frac{\hat{\gamma}_{H_0}^{(u)} \hat{\kappa}_{H_1}^{(u)} - \hat{\gamma}_{H_1}^{(u)} \hat{\kappa}_{H_0}^{(u)} + \sqrt{\hat{\gamma}_{H_0}^{(u)} \hat{\gamma}_{H_1}^{(u)} \hat{s}^{(u)}}}{\hat{\gamma}_{H_0}^{(u)} - \hat{\gamma}_{H_1}^{(u)}}.$$

Furthermore, $\hat{R}^{(u)}(\cdot)$ is twice differentiable at t_1^* and t_2^* , $\frac{d^2 \hat{R}^{(u)}}{dt^2}(t_1^*) > 0$ and $\frac{d^2 \hat{R}^{(u)}}{dt^2}(t_2^*) < 0$. Hence, by the second derivative test [56], t_1^* is a local minimum of $\hat{R}^{(u)}(\cdot)$, and t_2^* is a local maximum of $\hat{R}^{(u)}(\cdot)$.

Therefore, by Fermat's Theorem [56] and the fact that $\hat{R}^{(u)}(\cdot)$ is differentiable at any $t \in \mathbb{R}$, we conclude that exactly one of the following statements is satisfied:

- a) t_1^* is a global minimum of $\hat{R}^{(u)}(\cdot)$.
- b) $L_{10}P_{H_0} = \lim_{t \rightarrow -\infty} \hat{R}^{(u)}(t) \leq \hat{R}^{(u)}(r)$ for all $r \in \mathbb{R}$.
- c) $L_{01}P_{H_1} = \lim_{t \rightarrow \infty} \hat{R}^{(u)}(t) \leq \hat{R}^{(u)}(r)$ for all $r \in \mathbb{R}$.

If in addition to assumption C-1, assumption C-2 is satisfied then statement a must hold. Now, if assumption C-2 is not satisfied then statement b or statement c must hold, which means that $\hat{R}^{(u)}(t) > \min(L_{10}P_{H_0}, L_{01}P_{H_1})$ for all $t \in \mathbb{R}$, i.e. t_1^* is not a global minimum. Furthermore, if assumption C-1 is not satisfied then $\hat{R}^{(u)}(\cdot)$ has no stationary points and, again, t_1^* is not a global minimum. ■

F. Tukey's bi-square M-estimator implementation:

The considered Tukey's bi-square M-estimator of location minimizes the following objective function

$$J_\rho(\mathbf{a}) \triangleq \sum_{n=1}^N \rho\left(\frac{\|\mathbf{X}_n - \mathbf{a}\|}{\hat{\sigma}}\right),$$

where $\rho(r) \triangleq 1 - \left(1 - \left(\frac{r}{c}\right)^2\right)^3 \mathbb{1}_{[0,c]}(|r|)$ is Tukey's bi-square loss function, c is tuning constant that controls the asymptotic relative efficiency (ARE) [53] of the estimate relative to the CRLB under nominal Gaussian distribution, and $\mathbb{1}_{[0,c]}(\cdot)$ denotes the indicator function of the closed interval $[0, c]$. The robust scale parameter estimate $\hat{\sigma} \triangleq \sqrt{\frac{1}{p} \sum_{k=1}^p \hat{\sigma}_{X_k}^2}$, where

$$\hat{\sigma}_{X_k}^2 = \gamma^2 \left[\left(\text{MAD} \left(\{\text{Re}(\mathbf{X}_{k,n})\}_{n=1}^N \right) \right)^2 + \left(\text{MAD} \left(\{\text{Im}(\mathbf{X}_{k,n})\}_{n=1}^N \right) \right)^2 \right],$$

$\gamma \triangleq 1/\text{erf}^{-1}(3/4)$, is a robust median absolute deviation (MAD) estimate of variance [52]. The constant γ ensures consistency of the scale estimate under normally distributed data [52]. The ARE of the considered Tukey bi-square M-estimator, defined as the ratio between the traces of the CRLB and the asymptotic MSE under Gaussian distribution, is given by

$$\text{ARE}(c) = \frac{\mathbb{E}^2 \left[\left(1 - \left(\frac{R}{c} \right)^2 \right) \left(\left(\frac{2}{p} + 1 \right) \left(\frac{R}{c} \right)^2 - 1 \right) \mathbb{1}_{[0,c]}(R) \right]}{\mathbb{E} \left[\left(1 - \left(\frac{R}{c} \right)^2 \right)^4 \frac{R^2}{p} \mathbb{1}_{[0,c]}(R) \right]},$$

where $\sqrt{2}R$ is a chi distributed random variable with $2p$ degrees of freedom. Using this formula, the parameter c was set to achieve ARE of 95% in all simulation examples. For the considered dimension $p = 10$ we obtained $c \approx 6.2$. By equating the gradient of the objective function $J_\rho(\mathbf{a})$ to zero, Tukey's bi-square M-estimator of location is the solution of the equation $\hat{\mathbf{a}} = \frac{\sum_{n=1}^N \mathbf{X}_n w(\mathbf{X}_n, \hat{\mathbf{a}})}{\sum_{n=1}^N w(\mathbf{X}_n, \hat{\mathbf{a}})}$, obtained by fixed-point iteration, where the weight function $w(\mathbf{X}_n, \mathbf{a}) \triangleq \left(1 - \frac{\|\mathbf{X}_n - \mathbf{a}\|^2}{c^2 \hat{\sigma}^2} \right)^2 \mathbb{1}_{[0,c]} \left(\frac{\|\mathbf{X}_n - \mathbf{a}\|}{\hat{\sigma}} \right)$. Here, the fixed-point iteration was initialized by the robust median estimator of location. The maximum number of iterations and the stopping criterion in the fixed-point iteration were set to 100 and $\|\hat{\mathbf{a}}_l - \hat{\mathbf{a}}_{l-1}\| / \|\hat{\mathbf{a}}_{l-1}\| < 10^{-6}$, respectively, where l denotes an iteration index. Notice that unlike Tukey's location estimator, the empirical MT-mean (11) does not involve iterative optimization.

REFERENCES

- [1] L. L. Scharf, *Statistical signal processing*. Addison-Wesley, 1991.
- [2] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 231, pp. 289–337, 1933.
- [3] V. Spokoiny and T. Dickhaus, *Basics of Modern Mathematical Statistics*. Springer, 2015.

- [4] M. Taniguchi and Y. Kakizawa, *Asymptotic theory of statistical inference for time series*. Springer, 2012.
- [5] J. Bai, “Likelihood ratio tests for multiple structural changes,” *Journal of Econometrics*, vol. 91, no. 2, pp. 299–323, 1999.
- [6] K. Sakiyama and M. Taniguchi, “Discriminant analysis for locally stationary processes,” *Journal of Multivariate Analysis*, vol. 90, no. 2, pp. 282–300, 2004.
- [7] I. A. Canay, “Inference for partially identified models: Large deviations optimality and bootstrap validity,” *Journal of Econometrics*, vol. 156, no. 2, pp. 408–425, 2010.
- [8] M. C. Beaulieu, J. M. Dufour, and L. Khalaf, “Finite sample multivariate tests of asset pricing models with coskewness,” *Computational Statistics & Data Analysis*, vol. 53, no. 6, pp. 2008–2021, 2009.
- [9] S. Kullback and R. A. Leibler, “On information and sufficiency,” *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.
- [10] J. Miller and J. Thomas, “Detectors for discrete-time signals in non-Gaussian noise,” *IEEE Transactions on Information Theory*, vol. 18, no. 2, pp. 241–250, 1972.
- [11] S. M. Kay, *Fundamentals of statistical signal processing: Detection theory*. Prentice hall, 1998.
- [12] M. K. Tsatsanis and G. B. Giannakis, “Object and texture classification using higher-order statistics,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 7, pp. 733–750, 1992.
- [13] M. Coulon, J. Y. Tournier, and A. Swami, “Detection and classification of spectrally equivalent processes using higher order statistics,” *IEEE Transactions on Signal Processing*, vol. 47, no. 12, pp. 3326–3335, 1999.
- [14] B. Porat, *Digital processing of random signals: theory and methods*. Prentice-Hall, 1994.
- [15] K. Todros and A. O. Hero, “On measure transformed canonical correlation analysis,” *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4570–4585, 2012.
- [16] K. Todros and A. O. Hero, “Measure transformed canonical correlation analysis with application to financial data,” in *Proceedings of Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 361–364, 2012.
- [17] K. Todros and A. O. Hero, “Robust multiple signal classification via probability measure transformation,” *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1156–1170, 2015.
- [18] K. Todros and A. O. Hero, “Robust measure transformed MUSIC for DOA estimation,” in *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4190–4194, 2014.
- [19] K. Todros and A. O. Hero, “Measure-transformed quasi maximum likelihood estimation with application to source localization,” in *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3462–3466, 2015.
- [20] K. Todros and A. O. Hero, “Measure-transformed quasi maximum likelihood estimation,” *Accepted with mandatory minor revisions in the IEEE Trans. Signal Processing*, 2016. Arxiv version: <https://arxiv.org/pdf/1511.00237.pdf>.
- [21] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: John Wiley & Sons, 1973.
- [22] F. J. Jaimes-Romero and D. Munoz-Rodriguez, “Generalized Bayesian hypothesis testing for cell coverage determination,” *IEEE Transactions on Vehicular Technology*, vol. 49, no. 4, pp. 1102–1109, 2000.
- [23] J. Sohn, N. S. Kim, and W. Sung, “A statistical model-based voice activity detection,” *IEEE Signal Processing Letters*, vol. 6, no. 1, pp. 1–3, 1999.
- [24] J. Ramirez, J. C. Segura, C. Benitez, L. Garcia, and A. Rubio, “Statistical voice activity detection using a multiple observation likelihood ratio test,” *IEEE Signal Processing Letters*, vol. 12, no. 10, pp. 689–692, 2005.
- [25] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Mullers, “Fisher discriminant analysis with kernels,” in *Proceedings of the 1999 IEEE Signal Processing Society Workshop.*, pp. 41–48, 1999.

- [26] S. Mika, A. J. Smola, and B. Schölkopf, "An improved training algorithm for kernel Fisher discriminants," in *Proceedings of AISTATS*, 2001.
- [27] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural computation*, vol. 12, no. 10, pp. 2385–2404, 2000.
- [28] E. Pekalska and B. Haasdonk, "Kernel discriminant analysis for positive definite and indefinite kernels," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 6, pp. 1017–1032, 2009.
- [29] A. Berlinet and C. Thomas-Agnan, *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media, 2011.
- [30] S. Kay, "Can detectability be improved by adding noise?," *IEEE Signal Processing Letters*, vol. 7, no. 1, pp. 8–10, 2000.
- [31] H. Chen, P. K. Varshney, S. M. Kay, and J. H. Michels, "Theory of the stochastic resonance effect in signal detection: Part I - fixed detectors," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3172–3184, 2007.
- [32] H. Chen and P. K. Varshney, "Theory of the stochastic resonance effect in signal detection: Part II - variable detectors," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5031–5041, 2008.
- [33] G. B. Folland, *Real Analysis*. John Wiley and Sons, 1984.
- [34] P. Schreier and L. L. Scharf, *Statistical signal processing of complex-valued data: the theory of improper and noncircular signals*. Cambridge University Press, 2010.
- [35] I. S. Dhillon and J. A. Tropp, "Matrix nearness problems with bregman divergences," *SIAM Journal on Matrix Analysis and Applications*, vol. 29, no. 4, pp. 1120–1146, 2007.
- [36] K. B. Athreya and S. N. Lahiri, *Measure theory and probability theory*. Springer Science & Business Media, 2006.
- [37] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust statistics: the approach based on influence functions*. John Wiley & Sons, 2011.
- [38] W. Feller, *An introduction to probability theory and its applications: volume I*, vol. 3. John Wiley & Sons London-New York-Sydney-Toronto, 1968.
- [39] K. Todros and A. O. Hero, "Measure-transformed quasi likelihood ratio test," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4259–4263, 2016.
- [40] N. Halay, K. Todros, and A. O. Hero, "Measure transformed quasi likelihood ratio test for Bayesian binary hypothesis testing," in *Proceedings of Workshop on Statistical Signal Processing (SSP)*, 2016.
- [41] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5597–5625, 2012.
- [42] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.
- [43] K. Gerlach, "Spatially distributed target detection in non-Gaussian clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 3, pp. 926–934, 1999.
- [44] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5597–5625, 2012.
- [45] S. Visuri, H. Oja, and V. Koivunen, "Subspace-based direction-of-arrival estimation using nonparametric statistics," *IEEE Transactions on Signal Processing*, vol. 49, no. 9, pp. 2060–2073, 2001.
- [46] G. H. Golub and C. F. V. Loan, *Matrix Computations, 3rd ed.* Baltimore, MD: Johns Hopkins, 1996.
- [47] A. Swami, "On some parameter estimation problems in alpha-stable processes," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, pp. 3541–3544 vol.5, 1997.

- [48] A. Swami and B. Sadler, "TDE, DOA and related parameter estimation problems in impulsive noise," in *Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics*, pp. 273–277, 1997.
- [49] A. Swami and B. M. Sadler, "On some detection and estimation problems in heavy-tailed noise," *Signal Processing*, vol. 82, no. 12, pp. 1829–1846, 2002.
- [50] P. C. Allaart, "Minimax risk inequalities for the location–parameter classification problem," *Journal of multivariate analysis*, vol. 66, no. 2, pp. 255–269, 1998.
- [51] X. Ning and G. Karypis, "The set classification problem and solution methods," in *2008 IEEE International Conference on Data Mining Workshops*, pp. 720–729, IEEE, 2008.
- [52] P. J. Huber, *Robust statistics*. Springer, 2011.
- [53] R. J. Serfling, *Approximation theorems of mathematical statistics*. John Wiley & Sons, 2009.
- [54] R. Patel and M. Toda, "Trace inequalities involving hermitian matrices," *Linear algebra and its applications*, vol. 23, pp. 13–20, 1979.
- [55] H. B. Mann and A. Wald, "On stochastic limit and order relationships," *The Annals of Mathematical Statistics*, vol. 14, no. 3, pp. 217–226, 1943.
- [56] H. Anton, *Calculus, 5th ed.* John Wiley & Sons, 1995.